

Summer 2023

## Creating a People-First Court Data Framework

Lauren Sudeall

Charlotte S. Alexander  
*Scheller College of Business, Georgia Institute of Technology*

Follow this and additional works at: <https://scholarship.law.vanderbilt.edu/faculty-publications>



Part of the [Civil Law Commons](#), [Courts Commons](#), and the [Criminal Law Commons](#)

---

### Recommended Citation

Lauren Sudeall and Charlotte S. Alexander, *Creating a People-First Court Data Framework*, 58 *Harvard Civil Rights-Civil Liberties Law Review*. 731 (2023)

Available at: <https://scholarship.law.vanderbilt.edu/faculty-publications/1395>

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Vanderbilt Law. It has been accepted for inclusion in Vanderbilt Law School Faculty Publications by an authorized administrator of Scholarship@Vanderbilt Law. For more information, please contact [mark.j.williams@vanderbilt.edu](mailto:mark.j.williams@vanderbilt.edu).



DATE DOWNLOADED: Tue Jan 23 10:00:02 2024

SOURCE: Content Downloaded from [HeinOnline](#)

Citations:

Please note: citations are provided as a general guideline. Users should consult their preferred citation format's style manual for proper citation formatting.

Bluebook 21st ed.

Charlotte S. Alexander & Lauren Sudeall, Creating a People-First Court Data Framework, 58 HARV. C.R.-C.L. L. REV. 731 (2023).

ALWD 7th ed.

Charlotte S. Alexander & Lauren Sudeall, Creating a People-First Court Data Framework, 58 Harv. C.R.-C.L. L. Rev. 731 (2023).

APA 7th ed.

Alexander, C. S., & Sudeall, Lauren. (2023). Creating people-first court data framework. Harvard Civil Rights-Civil Liberties Law Review, 58(2), 731-788.

Chicago 17th ed.

Charlotte S. Alexander; Lauren Sudeall, "Creating a People-First Court Data Framework," Harvard Civil Rights-Civil Liberties Law Review 58, no. 2 (Summer 2023): 731-788

McGill Guide 9th ed.

Charlotte S. Alexander & Lauren Sudeall, "Creating a People-First Court Data Framework" (2023) 58:2 Harv CR-CL L Rev 731.

AGLC 4th ed.

Charlotte S. Alexander and Lauren Sudeall, 'Creating a People-First Court Data Framework' (2023) 58(2) Harvard Civil Rights-Civil Liberties Law Review 731

MLA 9th ed.

Alexander, Charlotte S., and Lauren Sudeall. "Creating a People-First Court Data Framework." Harvard Civil Liberties Law Review, vol. 58, no. 2, Summer 2023, pp. 731-788. HeinOnline.

OSCOLA 4th ed.

Charlotte S. Alexander & Lauren Sudeall, 'Creating a People-First Court Data Framework' (2023) 58 Harv CR-CL L Rev 731 Please note: citations are provided as a general guideline. Users should consult their preferred citation format's style manual for proper citation formatting.

Provided by:

Vanderbilt University Law School

-- Your use of this HeinOnline PDF indicates your acceptance of HeinOnline's Terms and Conditions of the license agreement available at

<https://heinonline.org/HOL/License>

-- The search text of this PDF is generated from uncorrected OCR text.

-- To obtain permission to use this article beyond the scope of your license, please use:

[Copyright Information](#)

# Creating a People-First Court Data Framework

Charlotte S. Alexander\* & Lauren Sudeall\*\*

*Most court data are maintained—and most empirical court research is conducted—from the institutional vantage point of the courts. Using the case as the common unit of measurement, data-driven court research typically focuses on metrics such as the size of court dockets, the speed of case processing, judicial decision-making within cases, and the frequency of case events occurring within or resulting from the court system.*

*This Article sets forth a methodological framework for reconceptualizing and restructuring court data as “people-first”—centered not on the perspective of courts as institutions but on the people who interact with the court system. We reorganize case-level data around the individual, identifying and analyzing the touchpoints that individuals have had over time with a range of different courts. In doing so, we invoke language as a signaling device to suggest a different, more intentional way to think about courts and the way we study their structure, processes, and impact.*

*The pilot research study that serves as the foundation for this Article is the first of its kind to apply a people-first approach to a data set that includes both criminal and civil state court records drawn from a random sample of 885 people in Fulton County, Georgia, between 2016 and 2020. Our methodology and findings provide a new perspective on the interactions between individuals and the courts and generate important new data relevant to a range of research areas. This approach and its results also represent a key step forward in expanding the application of a people-first approach to decentralized court systems, including those at the state and local levels. In taking this step, we empower and encourage researchers and policymakers at all levels to center those who experience the impact of court systems rather than focusing exclusively on the systems themselves.*

INTRODUCTION . . . . .	732
I. COURTS-FIRST VS. PEOPLE-FIRST DATA . . . . .	735
A. <i>Traditional Court Data Research: Courts-First</i> . . . . .	735
B. <i>The Importance of a People-First Approach</i> . . . . .	739
1. <i>Why People-First?</i> . . . . .	740
2. <i>Applying People-First Data to Court Research</i> . . . . .	742
II. PILOT STUDY METHODOLOGY . . . . .	748
A. <i>Pilot Study Site</i> . . . . .	749
B. <i>Criminal Case Number Assembly</i> . . . . .	750
C. <i>Defendant Name Lookups</i> . . . . .	753

---

\* Professor of Law and Ethics, Scheller College of Business, Georgia Institute of Technology. The authors are grateful to Mike Lissner, J.J. Prescott, Erika Rickard, Tanina Rostain, Colleen Shanahan, Jeffrey Vagle, and Nate Vogel for their thoughtful comments and feedback. For their outstanding research assistance, the authors thank Rahwa Asghedom, Jacqueline Avila, Clinton Cheang, Madison Gibbs, Rebecca Rhym, Zoe Siepert, and Maggie Willingham. This project received funding via the 2021 Open Justice Research Workshop seed grant competition, sponsored by the NSF-funded Systematic Content Analysis of Litigation Events Open Knowledge Network (SCALES OKN).

\*\* Professor of Law and Director, Vanderbilt Access to Justice Initiative, Vanderbilt Law School.

D. <i>Name Disambiguation</i> .....	756
III. FINDINGS .....	759
A. <i>What Do “People-First” Data Look Like?</i> .....	759
B. <i>Research Impact of People-First Data</i> .....	764
1. <i>Collateral Consequences</i> .....	765
2. <i>Holistic Legal Services and Problem-Solving Courts</i> .....	769
3. <i>Eviction</i> .....	770
IV. RESTRUCTURING COURT DATA .....	774
A. <i>Data Organization and Access Reforms</i> .....	775
B. <i>Protecting Privacy and Preventing Misuse</i> .....	778
1. <i>Backward-Looking Concerns</i> .....	779
2. <i>Forward-Looking Concerns</i> .....	781
CONCLUSION .....	784
APPENDIX A .....	786
APPENDIX B .....	788

## INTRODUCTION

Most court data are maintained—and most empirical court research is conducted—from the perspective of the courts as institutions. This dictates the way court data are structured and analyzed and encourages research that centers the courts themselves. For example, court data are commonly organized by the type of court from which they originate—state or federal, magistrate or superior, and civil or criminal. The common unit of measurement used in all these forums is the court case—a single dispute, controversy, or legal action taking place between a single set of parties.

Much court-focused research, therefore, is framed in terms of size of court dockets or caseloads, the speed with which courts process cases, and the types of case events that occur within or result from the court system. Other research builds on this model, directing more in-depth study to events occurring within cases, primarily judges’ decisions. This approach to court data—which we refer to as “courts-first”—is a result not only of how data are currently organized and what data are currently made available, but also choices made by court administrators and researchers about how we think about and study courts, their role, and the legal and social problems that come before them. Using the court case as the primary unit of analysis relies on a set of assumptions about how we conceive of and treat those problems—for example, what constitutes a “case” and how various legal issues relate to one another—as well as the target audience(s) for information generated by court research.

Here, we set out to interrogate some of those assumptions by reconceptualizing court data and creating a different picture of court activity, taken not from the vantage point of the courts but that of the people who experience the court system’s impacts. We center the individual as the relevant

locus of study,<sup>1</sup> identifying and analyzing the touchpoints that individuals have with a range of courts over time. Under this approach, case-centered analyses are not irrelevant, but are organized based on, and remain secondary to, the individual.

Rather than fragmenting the individual experience into distinct information silos—e.g., involvement in one court or one type of case—this approach attempts to capture a broader view of the whole person’s experience within and across the larger court system(s). We call this approach to court data “people-first,” given the primacy of the individual not only in collecting and organizing court data, but also in how those data are used and the story they tell about the courts’ operation.

We have drawn the “people-first” terminology from literature rooted in the belief that the way in which language is used—particularly in the legal context—can have powerful implications.<sup>2</sup> Person-first language centers the person rather than one characteristic or affiliation; in doing so, it recognizes that people are not defined by any one experience or aspect of their identity.<sup>3</sup> For example, a person-first approach would describe someone as a “person who has been convicted of a felony” rather than referring to that person as a “felon.”<sup>4</sup> Language can affect our perceptions of others, imbue them with certain characteristics or common understandings, and institutionalize those beliefs as part of broader systems.<sup>5</sup> It is for that reason that many groups, including disability rights<sup>6</sup> and criminal justice advocates,<sup>7</sup> have argued for the adoption of person-first language in their respective fields. This approach has since been adopted by scholars working in the institutional context, driving the use of similar terminology in statutory law.<sup>8</sup>

Like others,<sup>9</sup> we acknowledge that language is only a first step in the systems change required to actually center people at the core of policy

---

<sup>1</sup> For purposes of this piece, “individual” refers to a natural person who serves in the role of litigant, either plaintiff or defendant, but is consistent in identity across various data sets and distinct from any co-litigants.

<sup>2</sup> See Meg E. Ziegler, *Disabling Language: Why Legal Terminology Should Comport With A Social Model of Disability*, 61 B.C. L. REV. 1183, 1185 (2020).

<sup>3</sup> See Nancy G. La Vigne, *People first: Changing the Way we Talk About Those Touched by the Criminal Justice System*, URB. INST. (Apr. 5, 2016), <https://www.urban.org/urban-wire/people-first-changing-way-we-talk-about-those-touched-criminal-justice-system> [<https://perma.cc/HV6H-7RND>].

<sup>4</sup> See, e.g., Bill Keller, *Inmate. Parolee. Felon. Discuss.*, THE MARSHALL PROJECT (Apr. 1, 2015), <https://www.themarshallproject.org/2015/04/01/inmate-parolee-felon-discuss#.I1NCNhAc0> [<https://perma.cc/NQ8R-Q3KJ>].

<sup>5</sup> See *id.* at 1206–10. See generally SALLY MCCONNELL-GINET, WORDS MATTER (2020) (exploring the ways in which language not only reflects but also establishes social identities).

<sup>6</sup> See Ziegler, *supra* note 2.

<sup>7</sup> See La Vigne, *supra*, note 3.

<sup>8</sup> See, e.g., Rosa’s Law, Pub. L. No. 111-256, 124 Stat. 2643 (2010) (codified as amended in scattered sections of the U.S.C.) (mandating that “mental retardation” be replaced with person-first language in numerous federal laws).

<sup>9</sup> See, e.g., Talisa J. Carter, *Person-First Language Is Not Enough*, URB. INST. (May 28, 2021), <https://www.urban.org/urban-wire/person-first-language-not-enough> [<https://perma.cc/ARW5-W2U6>].

choices and development. Here, we aim to use language as a signaling device, to suggest a different way of thinking about courts and studying their processes and impact.<sup>10</sup> Just as person-first language encourages us to see a person as more than just one of her experiences, a people-first approach to court data enables us to understand the individual's experience with the court system as more than one isolated touchpoint. To that end, we offer a methodology for restructuring court data to facilitate that process and generate empirical research findings that can lead to systems change where it is needed. In doing so, our use of "people-first" is not only symbolic but also operational.

In advocating for court data models that allow for the application of a people-first approach, we do not mean to imply that one type of data—courts-first or people-first—is better or more necessary than the other. The relevance of data, and the vantage point that will be most useful, depends in large part on the research question being asked. For example, if the research question relates to the relative speed of case processing in urban and rural areas, court-level data will be the focus. If researchers are interested in judicial discrimination, they will need judge-level data. And if one is interested in how eviction affects individual employment and education outcomes, people-first data will be required. In some cases, people-first data tell just one part of an important larger story: for example, focusing only on person-level interactions with law enforcement in the context of a criminal case may hide larger patterns of discrimination or violence that can only be fully understood in the aggregate. Both courts-first and people-first court data are important to understanding how the courts do and should work; yet, at present, we have far greater ability to access the former than the latter.

In this Article, we present a pilot study on the generation and use of people-first data in a single set of courts: Fulton County, Georgia's Magistrate, State, and Superior courts. Ultimately, this pilot project is an attempt to restructure court data through a people-first orientation and explore the ramifications of that restructuring for the individuals involved, for courts, and for the researchers who study them.

In Part I, we describe in more detail the current "courts-first" approach to court data research. We then set forth and contrast a "people-first" approach to court data and elaborate the importance of such a shift. Last, we explain how a range of existing research areas—including collateral consequences, holistic legal services, problem-solving courts, and eviction—

---

<sup>10</sup> Neel Sukhatme's recent work on how imprisonment affects families of the convicted provides a striking view of the potential impact of person-centered research. Using birth records, Sukhatme links sentencing data with comprehensive information on criminal, health, voting, and economic outcomes across generations to better understand how incarceration affects not only individuals, but their loved ones. In doing so, he has shifted the structural frame for the analysis to adopt not only the individual's perspective, but those who are intimately connected to the individuals as well. See Neel Sukhatme, *The Impact of Criminal Sanctions on the Social and Economic Fabric of Families*, <https://www.carnegie.org/awards/honoree/neel-sukhatme/> [<https://perma.cc/G8LE-L4NZ>].

would benefit from a people-first court data approach. In Part II, we describe the methodological elements critical to our pilot people-first court data analysis and those necessary to define and generate people-first data from a courts-first data system. Part III delves into the findings from our pilot study, demonstrating the value that can be gleaned from people-first data and, by way of example, the relevance of those data to the research areas outlined above. In Part IV, we explore the requirements of a more systemic people-first restructuring of court data and some of the concerns that might accompany that shift, relating primarily to individual privacy and predictive use.

In critiquing the current state of data-driven approaches to legal reform, legal scholar Erin Collins recently advocated, “[W]e must expand our data collection and analysis practices to include and prioritize those that center the insight and goals of those who have experienced the violence and injustices of the system.”<sup>11</sup> Our proposed people-first model centers and advances that very notion, shifting the lens from the institutions that purport to do justice to those on whom (in)justice is done.

## I. COURTS-FIRST VS. PEOPLE-FIRST DATA

This Part first describes the current state of courts-first empirical scholarship and then introduces and contrasts the people-first data framework. It subsequently identifies four existing research areas that would benefit from people-first data and analysis: collateral consequences, holistic legal services, problem-solving courts, and eviction.

### A. *Traditional Court Data Research: Courts-First*

Much of the empirical scholarship on the U.S. judicial system takes as its site of study a particular court or courts, or a judge or judges, or those courts’ or judges’ handling of a particular type of case. The focus of this work may be on the courts themselves as actors, asking how different levels of the court system signal to and interact with one another,<sup>12</sup> how courts

---

<sup>11</sup> Collins argues: “This notion is not radical. In fact, the original evidence-based approach to medicine embraced a ‘bottom-up approach’ that considered insight from the needs and choices of the patient who was receiving treatment. Nor is it unrealistic; community-based participatory research methods were created to achieve these very aims. Meanwhile, we should heed the advice of QuantCrit scholars and embrace an attitude of ‘principled ambivalence’ towards the quantitative data we do have. We need not categorically reject statistical data—but to the extent we engage with such data, we should do so with an awareness of how it represents the product of a series of choices by those empowered to set the research agenda and conduct the studies—and on how those choices impact the outcome.” Erin Collins, *The Evidence-Based Trap*, INQUEST (Jul. 22, 2022), <https://inquest.org/the-evidence-based-trap/> [<https://perma.cc/7FKJ-W3ZP>].

<sup>12</sup> See, e.g., Tom S. Clark & Jonathan P. Kestellec, *The Supreme Court and Percolation in the Lower Courts: An Optimal Stopping Model*, 75 J. POL. 150, 151–52 (2013) (proposing the optimal intervention point for an appellate court given doctrinal conflict among lower courts); see also Andrew F. Daughety & Jennifer F. Reinganum, *Stampede to Judgement: Persuasive*

respond to rule changes about case processing times,<sup>13</sup> or how different courts vary in their implementation of the same procedural requirements.<sup>14</sup> Other branches of this work center on the decision-making of court officials, and investigate the relationship between judges' political ideologies or demographic characteristics and their rulings,<sup>15</sup> or the role of various types of judicial actors in handling cases.<sup>16</sup> Still other threads of this literature select certain case types and trace their path through the courts, by identifying judicial intervention points and attempting to explain or predict outcomes.<sup>17</sup>

Taken as a whole, this scholarship adopts a top-down, court's eye view of the administration of justice. Further, much of this work has focused even more narrowly on the extreme top of the judicial system. The U.S. Supreme Court, its justices, and its caseload—representing a tiny fraction of the busi-ness of this country's courts—have received an outside share of court-focused scholarly attention,<sup>18</sup> followed by the federal appellate courts.<sup>19</sup>

---

*Influence and Herding Behavior by Courts*, 1 AM. L. & ECON. REV. 158 (1999) (modeling courts' reliance on precedent as herding behavior).

<sup>13</sup> Miguel F. P. de Figueiredo, Alexandra D. Lahav, & Peter Siegelman, *The Six-Month List and the Unintended Consequences of Judicial Accountability*, 105 CORNELL L. REV. 363, 363 (2020) (analyzing the effect of the "six month list," which requires federal district court judges to make public reports of motions and cases pending in their courts twice a year).

<sup>14</sup> See, e.g., Adam R. Pah, David L. Schwartz, Sarath Sanga, Zachary D. Clopton, Peter Dicola, Rachel D. Mersey, Charlotte S. Alexander, Kristian J. Hammond, & Luís A. Nunes Amaral, *How to Build a More Open Justice System*, 369 SCIENCE 134, 135 (July 10, 2020), <https://www.science.org/doi/10.1126/science.aba6914/> [<https://perma.cc/X8GT-KANQ>] (finding differing filing fee waiver rates within and among U.S. district courts, where fee waiver decisions are governed by the same set of open-ended standards across courts).

<sup>15</sup> See generally Adam Bonica & Maya Sen, *Estimating Judicial Ideology*, 35 J. ECON. PERSP. 97 (2021) (summarizing literature on judicial ideology); Christina L. Boyd, *Representation on the Courts? The Effect of Trial Judges' Sex and Race*, 69 POL. RES. Q. 788 (2016) (studying judges' demographics and decision-making).

<sup>16</sup> See, e.g., Charlotte S. Alexander, Nathan Dahlberg, & Anne M. Tucker, *The Shadow Judiciary*, 39 REV. OF LITIG. 303 (2020) (studying interplay between magistrate judges' recommendations and U.S. district court judges' ultimate decisions); Christina L. Boyd & Jacqueline M. Sievert, *Unaccountable Justice? The Decision Making of Magistrate Judges in the Federal District Courts*, 34 JUST. SYST. J. 249 (2013) (studying magistrate judges' decision-making in selected U.S. district courts).

<sup>17</sup> See, e.g., Charlotte S. Alexander, *#MeToo and the Litigation Funnel*, 23 EMP. RTS. & EMPL. POL'Y J. 17 (2019) (studying sexual harassment case outcomes in a U.S. district court and summarizing previous literature on employment discrimination litigation outcomes); Christopher A. Cotropia, Jay P. Kesan & David L. Schwartz, *Heterogeneity Among Patent Plaintiffs: An Empirical Analysis of Patent Case Progression, Settlement, and Adjudication*, 15 J. EMP. LEGAL STUD. 80 (2018) (examining the relationship between plaintiff type and litigation outcomes in patent infringement lawsuits).

<sup>18</sup> See generally JEFFREY A. SEGAL & HOWARD J. SPAETH, *THE SUPREME COURT AND THE ATTITUDINAL MODEL REVISITED* (2002); FORREST MALTZMAN, JAMES F. SPRIGGS, & PAUL J. WAHLBECK, *CRAFTING LAW ON THE SUPREME COURT* (2000).

<sup>19</sup> For example, there are scholar-curated and maintained databases of U.S. Supreme Court and U.S. Court of Appeals opinions, but no equivalent for the trial-level U.S. district courts or, comprehensively, the state courts systems. See, e.g., HAROLD J. SPAETH, LEE EPSTEIN, ANDREW D. MARTIN, JEFFREY A. SEGAL, THEODORE J. RUGER & SARA C. BENESH, *THE SUPREME COURT DATABASE*, 2022 RELEASE 01, <http://scdb.wustl.edu/> [<https://perma.cc/H3K8-S44W>] (last visited June 14, 2022) ("The Supreme Court Database is the definitive source for researchers, students, journalists, and citizens interested in the U.S. Supreme Court."); THE SONGER PROJECT, *U.S. COURT OF APPEALS DATABASE*,



This is partly a problem of data availability. While many jurisdictions and private vendors make docket sheets, party-filed documents, and judges' decisions available electronically at differing costs, comprehensive, reliable, and freely available structured data sets on litigation do not exist.<sup>20</sup> In other words, courts are a text-rich but data- and information-poor environment.<sup>21</sup> With respect to the civil justice system in particular, comprehensive structured court data sets are especially difficult to create, given the multitude of ways in which data are tracked and stored (including in paper form), disaggregated case management systems, the lack of common terminology across courts, the lack of litigant demographic information, legal barriers to bulk electronic court records, and institutional barriers to making such data available.<sup>22</sup> As a result, it is difficult to capture the full, granular picture of litigants' regular interactions with the courts, including, for example, the frequency with which litigants are represented by attorneys and, if so, whether those attorneys are retained or appointed; the extent to which litigants appear repeatedly before the same judge; and patterns in the types of cases in which litigants are involved, including the pathways down which those cases proceed and their outcomes.

Data scarcity exerts substantial influence over which questions get studied, much as a person searching for lost keys at night limits their search to the area under a lamppost. In this instance, the lamp shines the most brightly on the top of the court system. It is easiest to study the Supreme Court because its docket is so small. Yet the information gained from tracking the Supreme Court's fewer than eighty average annual cases, logging justices' votes, and performing analyses is trivial compared to the over 350,000 new cases filed in the trial-level U.S. district courts per year, not to mention the multiple other levels of each state's court system.<sup>23</sup>

---

courts-of-appeals-databases.html [https://perma.cc/FM5J-FP8G] (last visited June 14, 2022) ("The Appeals Courts Database Project was designed to create an extensive dataset to facilitate the empirical analysis of the votes of judges and the decisions of the U.S. Courts of Appeals.").

<sup>20</sup> See Charlotte S. Alexander & Mohammad Javad Feizollahi, *On Dragons, Caves, Teeth, and Claws: Legal Analytics and the Problem of Court Data Access*, in COMPUTATIONAL LEGAL STUDIES: THE PROMISE AND CHALLENGE OF DATA-DRIVEN RESEARCH 95 (Ryan Whalen ed., 2020) (describing paucity of court data).

<sup>21</sup> This is a play on the acronym "DRIP," or "data-rich, information-poor," a term that originated in the 1980s business strategy literature. See generally THOMAS J. PETERS & ROBERT H. WATERMAN, JR., IN SEARCH OF EXCELLENCE: LESSONS FROM AMERICA'S BEST-RUN COMPANIES (1982) (originating the term). Put differently, the existence of extensive data in and of itself does not necessarily yield meaningful information, knowledge, or insight. Courts might be described as "TRIP," text-rich, information-poor, or less melodiously, "TRDIP," or "text-rich, data- and information-poor."

<sup>22</sup> See Tanina Rostain and Amy O'Hara, *The Civil Justice Data Gap*, in LEGAL TECH AND THE FUTURE OF CIVIL JUSTICE 493–99 (David Engstrom, ed. Cambridge University Press, forthcoming 2023); Claire Johnson Raba, *Low-Income Litigants in the Sandbox: Court Record Data and the Legal Technology Market*, ST. JOHN'S L. REV. (forthcoming).

<sup>23</sup> See LAWRENCE BAUM, THE SUPREME COURT 1 (14th ed. 2022) ("In the current era, the Supreme Court reaches full decisions in an average of fewer than eighty cases a year."); Christina L. Boyd, Pauline T. Kim & Margo Schlanger, *Mapping the Iceberg: The Impact of*

Even where scholars have engaged in time- and resource-intensive efforts to investigate how courts operate on the ground at lower levels of the system, much of that research has focused on the behavior of judges or observation of court processes.<sup>24</sup> Where research has explored the litigant experience in more depth, that story, too, typically focuses on courts and how the litigant's experience informs a particular court process (and its ability to achieve justice).<sup>25</sup> Thus, even this work tends to retain a systemic or institutionally-focused, courts-first lens.

Although scholarship focused more exclusively on the individual's experience with the legal system certainly exists, it is often of a qualitative or ethnographic nature<sup>26</sup> or focused primarily on one type of legal engagement<sup>27</sup>—in part, we suggest, because court data are not configured to easily support a multi-faceted quantitative approach to those inquiries. And where researchers attempt to connect people's experiences across case types and to other outside records, securing and reconciling data from all relevant sources can pose significant challenges.<sup>28</sup> Overcoming these obstacles and facilitating greater access to people-first court data can provide a critical quantitative element to the broader field of people-first research, which would allow for powerful mixed method research studies. It can also serve as an important foundation for creating linkages between court data sets and other sources of data that can shed light on the reasons for court involvement and its effects, including those relating to health, financial, and housing security.<sup>29</sup>

*Data Sources on the Study of District Courts*, 17 J. EMP. LEGAL STUD. 466, 466 (2020) (“Today, [federal district] courts receive well over 350,000 new civil and criminal cases per year, compared to just 50,000 matters in the federal courts of appeals and fewer than 80 merits cases at the U.S. Supreme Court annually.”).

<sup>24</sup> See, e.g., Anna E. Carpenter, Colleen F. Shanahan, Jessica K. Steinberg, & Alyx Mark, *Judges in Lawyerless Courts*, 110 GEO. L.J. 509 (2022) (exploring judicial behavior in *pro se* courts); Anna E. Carpenter, Jessica K. Steinberg, Colleen F. Shanahan, & Alyx Mark, *Studying the “New” Civil Judges*, 2018 WIS. L. REV. 249, 256–67 (describing research on the day-to-day operations of state civil courts, with a focus on the judicial role).

<sup>25</sup> See, e.g., Barbara Bezdek, *Silence in the Court: Participation and Subordination of Poor Tenants’ Voices in Legal Process*, 20 HOFSTRA L. REV. 533 (1992).

<sup>26</sup> See, e.g., Sara Sternberg Greene, *Race, Class, and Access to Civil Justice*, 101 IOWA L. REV. 1263 (2016); Kay Levine & Virginia Mellema, *Strategizing the Street: How Law Matters in the Lives of Women in the Street-Level Drug Economy*, 26 LAW & SOC. INQUIRY 169 (2001). See generally PATRICIA EWICK & SUSAN S. SIBLEY, *THE COMMON PLACE OF LAW: STORIES FROM EVERYDAY LIFE* (1998).

<sup>27</sup> This includes, for example, studies of recidivism, which explore individuals’ repeated contacts with the criminal legal system. See, e.g., James M. Anderson, Maya Buenaventura, and Paul Heaton, *The Effects of Holistic Defense on Criminal Justice Outcomes*, 132 HARV. L. REV. 819, 823 (2019) (exploring the effects of holistic defense model on case outcomes and future interactions with the criminal legal system).

<sup>28</sup> See, e.g., J.J. Prescott and Sonja B. Starr, *Expungement of Criminal Convictions: An Empirical Study*, 133 HARV. L. REV. 2460, 2465–66 (2020) (describing an unprecedented statewide study” made possible by a data-sharing agreement with the State of Michigan, that linked de-identified criminal record histories with wage and employment data to study the effects of expungement).

<sup>29</sup> Rostain and O’Hara, *supra* note 22, at 488–89 (highlighting research avenues that “focus on the people involved, not the court encounters themselves”).

As legal empiricists Kevin Clermont and Ted Eisenberg have argued, the paucity and limitations of existing court data “restrict[ ] what one can study about the legal system, and surely make[ ] risky any behavioral inferences one might draw therefrom.”<sup>30</sup> Legal scholar Lynn LoPucki has observed further, “By offering selective access to data, the courts have controlled legal scholars’ research agendas, . . . discouraging research that focused on the actions of judges and the impacts of those actions on both litigants and the public.”<sup>31</sup> Tanina Rostain and Amy O’Hara have described in detail the barriers to data access that prevent researchers from achieving a full understanding of “the consequences of court involvement for the life cycle of poverty across sites and population groups.”<sup>32</sup> In addition to issues of data access, the way in which data are collected by courts—and the information they choose to capture or include—can make it difficult to construct a complete picture of the justice ecosystem.

Our intervention shines light where much of existing research has not been able to fully explore: the impact on and experiences of the people on whom, and on whose behalf, the courts act. In the next section, we propose a people-first lens. Many of the data access problems hindering some strands of courts-first scholarship complicate our people-first approach as well. Yet the people-first pilot study we present here provides evidence of the promise of this approach.

### B. *The Importance of a People-First Approach*

In its simplest form, people-first research is research that centers the individual litigant as the focus of study. This contrasts with much of court data research, which focuses on the courts themselves as institutional actors or the officials who carry out court business, such as judges. Much of courts-first research uses the *case* as the sole unit of analysis. While people-first research might also encompass the study of cases, the cases are clustered according to the *people* involved. Alternatively, a people-first approach might focus only on the *people* interacting with the courts, without regard to their cases at all.<sup>33</sup>

More broadly, people-first research can be understood as encompassing both a *method of data collection* (individual person as focus) and also a *means or manner of reporting data* (from the individual’s perspective).

---

<sup>30</sup> Kevin M. Clermont & Theodore Eisenberg, *Litigation Realities*, 88 CORNELL L. REV. 119, 129 (2002).

<sup>31</sup> Lynn M. LoPucki, *The Politics of Research Access to Federal Court Data*, 80 TEX. L. REV. 2161, 2162 (2002).

<sup>32</sup> Rostain and O’Hara, *supra* note 22, at 492; *id.* at 482 (“The lack of high-quality, accessible data is a major deterrent to producing knowledge about civil justice.”). Rostain and O’Hara also describe the newly established Civil Justice Data Commons, housed at Georgetown University, which aims to provide a “regulated marketplace” to share civil justice data as well as related research tools and methods. *Id.* at 499–501.

<sup>33</sup> We do both in presenting our findings in Part III below.

Through that lens, people-first research is not necessarily a binary construct but operates along a spectrum. For example, quantifying the number of people who interact with a court each year would be further along the people-first spectrum than reporting the number of cases processed by a court each year. But further yet along the spectrum is an analysis that examines the process from the individual's perspective (asking, for example, how many courts the individual interacted with in one year or five years) rather than from the perspective of the court (inquiring how many cases the court processed during those same periods).

In this section, we discuss the importance of adopting a people-first approach to court data collection and research. First, we explain how people-first court data can help establish a foundational descriptive picture of the individual court experience and provide important new insights. Second, we demonstrate how this approach could expand the scope of existing research, address outstanding questions, and allow new questions to be posed, providing several examples of substantive research areas that could benefit from such data.

### *1. Why People-First?*

The traditional top-down, court-focused lens described in Part I.A provides an important perspective on how court systems operate—but also an incomplete one. Using only a courts-first lens, there is a tendency to analyze system outcomes according to the variables and metrics typically utilized by institutions: speed, volume, and dispositions obtained in one specific type of court. Because courts-first research uses cases as its unit of analysis, it can be difficult to disaggregate and track the experience of individual litigants. For example, where multiple parties are grouped together in a single case, case-level analyses elide differences between individuals' litigation pathways and outcomes. Moreover, there is little reason to consider how the processes and timeline in one court—as experienced by a given individual—relate to that individual's experience with other courts, or about the relationship between individual outcomes in one court and the initiation or conclusion of proceedings involving the same individual in another. Courts-first research also lends itself to siloing: because the court system itself is broken into discrete categories—criminal and civil, state and federal—court research tends to track similar lines. Yet any attempt at developing a complete understanding of court systems and structures requires a broader, more inclusive data gathering process.

Lynne Haney's recent work on incarcerated fathers provides one example of the importance of reexamining court systems from the perspective of the individual. Haney has observed that while there is significant research on both the child support and criminal justice systems, we know relatively little

about the connections and overlaps between the two.<sup>34</sup> For incarcerated fathers, Haney writes, entanglements between the two systems are “multidirectional” and “work in circular ways to form feedback loops of disadvantage.”<sup>35</sup> Haney explains how each system’s insistence on its distinct nature runs counter to the individual’s experience and ultimately disserves those who are subjected to them:

While child support and criminal justice institutions might insist on their separateness, parents experience their interconnections. Indeed, part of the power of these state systems lies in their *denial* of those intersections. Each system has its own separate demands and expectations, yet parents living between the two of them encounter these demands as crisscrossing. From setting support orders *in absentia* to the legal processes of child support court, these systems feed off each other. This makes incarcerated parents’ relationship to child support different from that of other low-income parents. Incarcerated parents are situated across institutions in ways that consistently trip them up[.]<sup>36</sup>

Gaining a deeper understanding of the difficulties incarcerated fathers face requires multiple research strategies, including those that incorporate factors and events occurring outside of the legal system. Yet an important part of clarifying the picture is understanding how and when these fathers interact with distinct parts of the court system, both civil and criminal.<sup>37</sup>

Further, as our pilot study findings presented in Part III.B suggest, a *child’s* involvement in a child support case may be associated with later civil and/or criminal legal involvement once the child reaches maturity. Though this link is only suggestive and far from probative, it points to another avenue for research like Haney’s, and perhaps resulting policy intervention, that could be enabled by people-first data.

Some scholars have already made headway in taking a more people-centered approach to court data research, corraling data from multiple sources through a significant investment of time and resources.<sup>38</sup> For exam-

---

<sup>34</sup> See Lynne Haney, *Making Men Pay*, INQUEST (Jun. 17, 2022), <https://inquest.org/making-men-pay/> [<https://perma.cc/Z7J5-TFJA>].

<sup>35</sup> *Id.*

<sup>36</sup> *Id.*

<sup>37</sup> See, e.g., Tonya L. Brito, David J. Pate, Jr., & Jia-Hui Stefanie Wong, “*I Do for My Kids*”: *Negotiating Race and Racial Inequality in Family Court*, 83 *FORDHAM L. REV.* 3027, 3033–36 (2015) (detailing the experiences of low-income non-custodial fathers in the child support context).

<sup>38</sup> For example, one study, discussed in more detail below, utilized a dyad-based linking system, linking two agency files at a time as a means for identifying files relating to the same individual. See Jessica A. Kelly, *Federal Justice Statistics Program Data Linking System*, URB. INST. (2012), <https://www.ojp.gov/library/publications/federal-justice-statistics-program-data-linking-system> [<https://perma.cc/4MR7-87ZK>]. This is more challenging in state court and law enforcement systems, which are typically more fragmented and may not share as many common variables. See *infra* Part IV.

ple, M. Merit Rehavi and Sonja Starr collected records from multiple agencies—federal law enforcement, prosecutors, courts, and the U.S. Sentencing Commission—to construct a picture of individual defendants’ paths through the federal criminal legal system.<sup>39</sup> Ultimately, their project focused on gaining a better understanding of disparities within the criminal court system, more so than elevating the individual experience. Yet there is room to push this approach further, breaking down civil-criminal barriers and adopting a more holistic approach.<sup>40</sup> Further, although there is a growing body of research exploring state civil courts through a people-focused lens, Anna Carpenter, Alyx Mark, Colleen Shanahan, and Jessica Steinberg have recently emphasized in surveying the literature that “we know a lot more about how people experience civil legal problems *outside* of the courthouse than we know about what happens *inside* the courthouse.”<sup>41</sup>

## 2. Applying People-First Data to Court Research

As described above, there is still much we do not know about how and when people interact with courts, particularly state courts.<sup>42</sup> As an initial matter, descriptive research about how, when, and in what capacity people interact with a range of different courts—both criminal and civil—would help researchers understand the contours of the ordinary person’s exposure to and experience with the courts. It would also be incredibly useful to the legal and nonlegal service providers that support those individuals to provide empirical data to complement their existing understanding of their clients’ experiences, and for consideration in structuring their service offerings.

Beyond that foundational descriptive work, there are several research areas that could specifically benefit from a court data set structured around the individual. In this section, we highlight four examples of substantive research areas in which people-first data can elucidate the current state of the issue, demonstrate potential for growth, answer outstanding questions, and help suggest directions for effective reform. We return to these topics in Part III.B, where we present examples from the findings of our people-first pilot study that are relevant to each of these four areas.

---

<sup>39</sup> See M. Marit Rehavi and Sonja B. Starr, *Racial Disparity in Federal Criminal Sentences*, 122 J. POL. ECON. 1320 (2014).

<sup>40</sup> In contrast to these studies, which establish connections across agencies and courts within the criminal justice system, our methodology reaches further, including all of the civil and criminal courts within one jurisdiction, Fulton County, Georgia. See *infra* Parts II.A and III.A. A further extension of our work could reach across all courts across all jurisdictions.

<sup>41</sup> Anna E. Carpenter, Alyx Mark, Colleen F. Shanahan & Jessica K. Steinberg, *Foreword: The Field of State Civil Courts*, 122 COLUM. L. REV. 1165, 1177 (2022).

<sup>42</sup> See Anna E. Carpenter, Jessica K. Steinberg, Colleen F. Shanahan, & Alyx Mark, *Studying the “New” Civil Judges*, 2018 WIS. L. REV. 249, 250 (2018); Stephen C. Yeazell, *Courting Ignorance: Why We Know So Little About Our Most Important Courts*, 143 DÆDALUS, J. AM. ACAD. ARTS & SCI. 129, 129 (2014).

a. *Collateral consequences*

Collateral consequences are traditionally understood as the range of effects that follow a criminal conviction, whether imposed formally by law or informally (extralegally) as a matter of practice.<sup>43</sup> Depending on the jurisdiction, collateral consequences may include the loss of the right to vote, serve on a jury, or hold office; loss of the ability to live in public housing or hold a driver's license; ineligibility for certain occupational licenses or types of employment; and the inability to maintain family relationships, including child custody and visitation.<sup>44</sup>

Much of the scholarly and popular attention on collateral consequences has focused on their prevalence and their wide range of effects.<sup>45</sup> Far more can be learned about the full nature of their impact, including the extent to which conviction is correlated with concurrent or future court involvement, whether criminal or civil. In addition, research about the nature and timing of subsequent court involvement can inform the debate about the extent to which collateral consequences can or do advance their purported purposes, including public safety,<sup>46</sup> and whether they might constitute “punishment” and therefore raise constitutional or procedural concerns.<sup>47</sup> Future court involvement may itself be a distinct category of collateral consequences, which person-first research could help define. Similarly, people-first data can help elucidate the cumulative effect of fines and fees imposed over time

---

<sup>43</sup> Wayne A. Logan, *Informal Collateral Consequences*, 88 WASH. L. REV. 1103, 1104 n.7 (2013) (explaining that formal collateral consequences include sanctions imposed by law as a result of conviction and discretionary disqualifications “imposed after an individualized inquiry by a legal authority”; informal collateral consequences, by contrast, “aris[e] independently of specific legal authority”).

<sup>44</sup> Gabriel J. Chin, *Collateral Consequences and Criminal Justice: Future Policy and Constitutional Directions*, 102 MARQ. L. REV. 233, 235 (2018).

<sup>45</sup> For example, the degree to which such consequences are imposed by law has been documented by various sources. See, e.g., *Compilations and Inventories, COLLATERAL CONSEQUENCES RES. CTR.*, <https://ccresourcecenter.org/compilations-inventories-of-collateral-consequences/> [<https://perma.cc/22SN-BCFW>]. Others have explored the range of economic, medical, psychological and social consequences that may not be formally imposed by law, but nonetheless result from incarceration—and may affect not only the person who has been incarcerated but also their family and friends. See Wayne A. Logan, *Informal Collateral Consequences*, 88 WASH. L. REV. 1103, 1104 (2013).

<sup>46</sup> See, e.g., MARGARET COLGATE LOVE, JENNY ROBERTS, & WAYNE A. LOGAN *COLLATERAL CONSEQUENCES OF CRIMINAL CONVICTION: LAW, POLICY, AND PRACTICE* 6 (2021–22 ed.) (asking whether collateral consequences serve “an important and legitimate safety purpose”); Gabriel J. Chin, *Collateral Consequences*, in 4 *REFORMING CRIMINAL JUSTICE* (2017) (“Collateral consequences should be rationalized and reformed to promote public safety, fairness in individual cases, and a more effective overall criminal justice system”); Jenny Roberts, *Gundy and the Civil-Criminal Divide*, 17 OHIO ST. J. CRIM. L. 207, 216–21 (2019) (demonstrating that a growing body of research has undermined the public safety rationale for many collateral consequences).

<sup>47</sup> See Lauren Sudeall, *Rethinking the Civil-Criminal Distinction*, in *TRANSFORMING CRIMINAL JUSTICE: AN EVIDENCE-BASED AGENDA FOR REFORM* (NYU Press, forthcoming); Brian M. Murray, *Are Collateral Consequences Deserved?*, 95 NOTRE DAME L. REV. 1031 (2020); Carol S. Steiker, *Punishment and Procedure: Punishment Theory and the Criminal-Civil Procedural Divide*, 85 GEO. L. J. 775 (1997).

across multiple areas (both criminal and civil) and their potential to redirect individuals back into the court system.

Construed more broadly, collateral consequences can also be understood as flowing from involvement with the civil legal system.<sup>48</sup> Kathryn Sabbeth explained that collateral consequences are not unique to the criminal context: civil court decisions can also result in a wide range of consequences, including a lack of access to housing, employment, and one's children.<sup>49</sup> Under that premise, a person-centered data approach would generate similar benefits. In addition, Rebecca Sandefur has emphasized that while we now have a fair amount of information about the civil legal needs of the U.S. public, large research gaps remain—including how civil legal needs affect the people who experience them.<sup>50</sup> While a large number of these impacts—economic, social, physiological, and psychological—will occur outside the courts, some of those impacts will be within the court system or reflected in those data. We can only see them, however, once the data are restructured to flow through the individual person.

*b. Holistic legal services*

The traditional model for public defense services focuses almost exclusively on criminal representation.<sup>51</sup> In contrast, the holistic legal services model recognizes that indigent clients “may be best served by a team of professionals that addresses a range of the client’s needs.”<sup>52</sup> Through the holistic model, therefore, defenders also address the collateral legal consequences of their clients’ criminal justice involvement and the underlying nonlegal issues that can lead to such involvement.<sup>53</sup> While this model has been most discussed and studied in the criminal context,<sup>54</sup> the concept of holistic legal services—based on the idea that any client, civil or criminal, might benefit from the same arrangement—may apply more broadly. To date, there has been little empirical research regarding the effectiveness of such a model; the first “rigorous, large-scale empirical evaluation of the holistic approach to indigent defense” occurred in 2019.<sup>55</sup> That study, like

---

<sup>48</sup> See Kathryn A. Sabbeth, *The Prioritization of Criminal Over Civil Counsel and the Discounted Danger of Private Power*, 42 FLA. ST. U. L. REV. 889, 913-14 (2015).

<sup>49</sup> See *id.*

<sup>50</sup> See generally Rebecca L. Sandefur, *What We Know and Need to Know About the Legal Needs of the Public*, 67 S.C. L. REV. 443 (2016).

<sup>51</sup> See James M. Anderson, Maya Buenaventura & Paul Heaton, *The Effects of Holistic Defense on Criminal Justice Outcomes*, 132 HARV. L. REV. 819, 821 (2019).

<sup>52</sup> *Id.*

<sup>53</sup> *Id.*

<sup>54</sup> See, e.g., Anderson et al., *supra* note 51; BRIAN J. OSTROM & JORDAN BOWMAN, EXAMINING THE EFFECTIVENESS OF INDIGENT DEFENSE TEAM SERVICES: A MULTISITE EVALUATION OF HOLISTIC DEFENSE IN PRACTICE, NATIONAL CENTER FOR STATE COURTS (2020); Sarah Buchanan & Roger M. Nooe, *Defining Social Work within Holistic Public Defense: Challenges and Implications for Practice*, 62 Social Work 333-39 (2017).

<sup>55</sup> Anderson et al., *supra* note 51, at 822. Anderson et al., like other researchers, measure recidivism in terms of rearrest(s) per person. *Id.* at n. 237. The present pilot study examines



others similar in nature, focused on the effects or outcomes of such a model, with specific attention to recidivism.<sup>56</sup>

Holistic legal needs can occur across the civil-criminal divide, but also across the many types of cases that can arise in just one sphere. Comparative studies have helped to shed light on patterns that occur across a wide range of justiciable problems, including family law issues, domestic violence, and consumer, employment, and debt problems.<sup>57</sup> In many of those studies, the incidence of justiciable problems was drawn from survey responses. Court data would provide a useful empirical complement.<sup>58</sup>

Better information about the nature of individuals' touchpoints across the full spectrum of civil and criminal courts can shed additional light on the possible effectiveness of holistic legal services, in both the civil and criminal settings. It can also provide critical information to determine what holistic services are most needed and when. For example, knowing that certain touchpoints tend to cluster together, happen in similar sequences, or with specific timing can help providers know what type of assistance is most critical at a certain point in time. Our results in Part III.B begin to suggest how such analyses can be performed within a people-first data framework.

*c. Problem-solving (specialized) and unified courts*

Problem-solving courts—sometimes known as specialized courts—have become increasingly popular as a form of criminal court reform.<sup>59</sup> These courts are criminal or quasi-criminal in nature and offer treatment and enhanced supervision in addition to, or in lieu of, incarceration.<sup>60</sup> Based on the premise that many interactions with the criminal legal system can be avoided through addressing underlying non-legal problems, problem-solving courts have emerged in a number of different areas—including mental health, drugs, veterans, homelessness, and domestic violence—at both the state and federal levels.<sup>61</sup> Although many of these courts claim to implement proven, “evidence-based” practices, data regarding their effectiveness are mixed<sup>62</sup> and many of these courts have not been thoroughly assessed, leaving

---

cases filed in court and does not include arrest data; we therefore do not make claims about “recidivism” as the concept is defined elsewhere in the literature.

<sup>56</sup> See *id.* See generally BRIAN J. OSTROM & JORDAN BOWMAN, EXAMINING THE EFFECTIVENESS OF INDIGENT DEFENSE TEAM SERVICES: A MULTISITE EVALUATION OF HOLISTIC DEFENSE IN PRACTICE, NATIONAL CENTER FOR STATE COURTS (2020).

<sup>57</sup> See generally NIGEL BALMER, ALEXY BUCK, & PASCOE PLEASANCE, CAUSES OF ACTION: CIVIL LAW AND SOCIAL JUSTICE (2006).

<sup>58</sup> *Id.*

<sup>59</sup> See Erin R. Collins, *The Problem of Problem-Solving Courts*, 54 U.C. DAVIS L. REV. 1573, 1576 (2021); Allegra M. McLeod, *Decarceration Courts: Possibilities and Perils of a Shifting Criminal Law*, 100 GEO. L.J. 1587, 1590 (2012).

<sup>60</sup> See Collins, *supra* note 59, at 1582.

<sup>61</sup> See Collins, *supra* note 59, at 1575–77.

<sup>62</sup> See, e.g., Collins, *supra* note 59, at 1578 (noting “mixed” results of drug court evaluations, showing some that decreased recidivism, some that increased recidivism, and yet others that had no impact on recidivism).

the full picture of their impact unclear.<sup>63</sup> Much of the empirical research on specialized courts attempts to measure whether they reduce costs or prevent recidivism.<sup>64</sup>

A person-centered approach would provide additional data to use in assessing problem-solving courts' effectiveness. Given their focus on criminal recidivism, many of these courts target interactions with the criminal legal system. Yet this model need not be limited to the criminal arena: Jessica Steinberg has written about the potential to extend the problem-solving court model to the civil context (beyond the family law context, where it already exists to some extent).<sup>65</sup> Additional empirical data on how, when, and in what capacity individuals in similar situations tend to interact with the civil courts could not only help to identify additional opportunities for application of the problem-solving court model, but also assist in structuring the types of assistance that existing problem-solving courts offer. For example, knowledge that a particular type of defendant (e.g., a veteran or person with a history of substance abuse) tends to have a certain pattern in their court interactions might help to flag *pro se* litigants for early intervention by specialized courts and enable court personnel to make necessary connections with other courts and service providers to resolve the full range of a defendant's current and future legal problems.<sup>66</sup>

In a similar vein, some family law scholars have advocated for unified family courts—courts that have jurisdiction to handle the wide range of matters that may arise in conjunction with children and families.<sup>67</sup> This has potential not only to reduce the number of hearings and judges required, but also to allow for more effective resolutions. For example, domestic violence issues may be intertwined with divorce and custody issues, criminal cases, dependency, and child support. Other issues may also be intertwined—such as the spectrum of issues that may arise from homelessness<sup>68</sup>—and could benefit from holistic treatment by the courts. As in the problem-solving court context, person-centered data can help identify the most commonly linked issues, any patterns in how they tend to arise, and questions of timing. They can also help inform how courts attempting to address such issues can be designed or structured to do so most effectively.

---

<sup>63</sup> Collins, *supra* note 59, at 1588 (noting that, “with the exception of drug courts, it is widely accepted that problem-solving courts have not been analyzed with rigor sufficient to form a conclusion about their impact”).

<sup>64</sup> See, e.g., McLeod, *supra* note 59, at 1591 n.12 (citing several studies).

<sup>65</sup> See Jessica K. Steinberg, *A Theory of Civil Problem-Solving Courts*, 93 N.Y.U. L. REV. 1579, 1584 (2018).

<sup>66</sup> Potential concerns about the predictive use of such data are addressed in Part IV.

<sup>67</sup> See, e.g., Catherine J. Ross, *The Failure of Fragmentation: The Promise of a System of Unified Family Courts*, 32 FAM. L. Q. 3 (1998).

<sup>68</sup> See Sara K. Rankin, *Hiding Homelessness: The Transcarceration of Homelessness*, 109 CALIF. L. REV. 559, 561–68 (2021).

d. *Eviction*

As housing costs and the rent burden on tenants have increased, more individuals and families across the U.S. are at higher risk of eviction.<sup>69</sup> While the eviction process is used by some landlords to reclaim property, in many other cases, it is used as a vehicle for rent collection.<sup>70</sup> This is true in the case of serial eviction filings, where a landlord or property owner files for eviction more than once against the same tenant, but stops the proceedings once the tenant pays the rent owed plus fines and fees. Serial filings are more common where there are fewer legal, regulatory, and economic barriers to filings and where the landlord is not an individual, but a corporation.<sup>71</sup>

In recent years, as attention on eviction has increased, more research on serial filing has emerged, focused on its frequency, its underlying motivations, and its effects on tenants within and beyond the housing context.<sup>72</sup> For example, Lillian Leung and her co-authors document the increased burden of fines and fees that tenants faced with serial eviction filings must pay, which landlords collect on top of back rent.<sup>73</sup> Further, serial eviction filings damage tenants' prospects of securing future housing.<sup>74</sup> Other research has highlighted the effects of housing displacement more generally, which include negative employment, education, medical, and housing related outcomes.<sup>75</sup>

There is still a research gap, however, with respect to how eviction filings and dispositions—whether individual or serial—impact the individual's broader legal or court-based experience. As is the case for many other substantive research areas,<sup>76</sup> eviction researchers have focused primarily on the eviction process itself,<sup>77</sup> and the extralegal inputs and outputs that affect it—including the housing market, landlord demographics and motivations, and the likelihood of securing new housing. For many tenants, however, their eviction experience may be one of many touchpoints with the courts,

---

<sup>69</sup> See Lillian Leung, Peter Hepburn, & Matthew Desmond, *Serial Eviction Filing: Civil Courts, Property Management, and the Threat of Displacement*, 100 *SOC. FORCES* 316, 317 (2021); see MATTHEW DESMOND, *EVICTED: POVERTY AND PROFIT IN THE AMERICAN CITY* (2016).

<sup>70</sup> See Philip ME Garboden & Eva Rosen, *Serial Filing: How Landlords Use the Threat of Eviction*, 18 *CITY & CMTY.* 638, 639–42 (2019).

<sup>71</sup> Leung et al., *supra* note 69, at 317.

<sup>72</sup> See, e.g., Garboden & Rosen, *supra* note 70, at 638, 656–57.

<sup>73</sup> See, e.g., Leung et al., *supra* note 69, at 317, 336.

<sup>74</sup> *Id.* at 317.

<sup>75</sup> See Kathryn A. Sabbeth, *Housing Defense as the New Gideon*, 41 *HARV. J. L. & GEN. DER.* 55, 66 (2018); Matthew Desmond, *Eviction and the Reproduction of Urban Poverty*, 118 *AM. J. SOC.* 88, 118 (2012). See generally Matthew Desmond & Rachel Tolbert Kimbro, *Eviction's Fallout: Housing, Hardship, and Health*, 94 *SOC. FORCES* 295 (2015).

<sup>76</sup> See, e.g., Sara Sternberg Greene, *Race, Class, and Access to Civil Justice*, 101 *IOWA L. REV.* 1263, 1290 (2016) (describing the siloed nature of legal research).

<sup>77</sup> See, e.g., Peter Hepburn, Renee Louis, & Matthew Desmond, *Racial and Gender Disparities among Evicted Americans*, 7 *SOC. SCI.* 649, 650 (2020) (relying on a data set composed only of eviction court records); Leung et al., *supra* note 69, at 320 (basing analysis on a data set composed exclusively of eviction cases); Garboden & Rosen, *supra* note 70 (same).

forming a larger, more textured story of how the state has shaped and continues to shape their lives. As above, we return to this problem in Part III.B below, using people-first data from our pilot study to investigate the prevalence and distribution of different patterns of eviction filings and the additional civil and criminal cases that eviction defendants may face.

As described above, there are many ways in which people-first court data can help researchers—as well as advocates and policymakers—better understand the effectiveness of court processes, court models, and legal services provision. Additionally, a people-first approach can aid courts in their regular operations and benefit court users. To provide just one example, family courts in Michigan adopt a “one judge, one family” approach, assigning family divisions jurisdiction over a wide range of matters, including name changes, parental consent for abortion, protection orders, and juvenile cases.<sup>78</sup> Under this model, families can ensure that cases in various areas are treated with consistency, in addition to saving time and resources and avoiding unnecessary confusion. Courts thus increase their efficiency by eliminating duplication and conserving judicial resources. Although such court models are not without their own limitations and risks,<sup>79</sup> the creation of people-first data could facilitate and support thoughtful and well-researched court restructuring efforts. More generally, it would provide courts with a valuable tool to assess their own effectiveness, evaluate how best to allocate their resources, and determine how well they serve the people who come before them. Thus, people-first data need not be viewed only as an academic endeavor; it can be just as useful to those engaged in the process itself.<sup>80</sup>

## II. PILOT STUDY METHODOLOGY

This Part turns to our pilot project’s methodology, with an emphasis on the aspects of the methodology most relevant to creating and collecting people-first data and the barriers we encountered. We narrate our workflow at a fairly granular level to illustrate the lengths required to generate even a small sample set of people-first court data in a pilot setting. We also offer this narration in the hopes that other researchers and court administrators will replicate, improve upon, and extend this work to other jurisdictions. Draw-

---

<sup>78</sup> Stacy Sellek, ‘One Judge, One Family’: Celebrating 20 Years of Michigan’s Family Division, SCAO CONNECTIONS (Jan. 12, 2018), <https://scao-connections.blogspot.com/2018/01/one-judge-one-family-celebrating-20.html> [<https://perma.cc/2J2V-5AL3>].

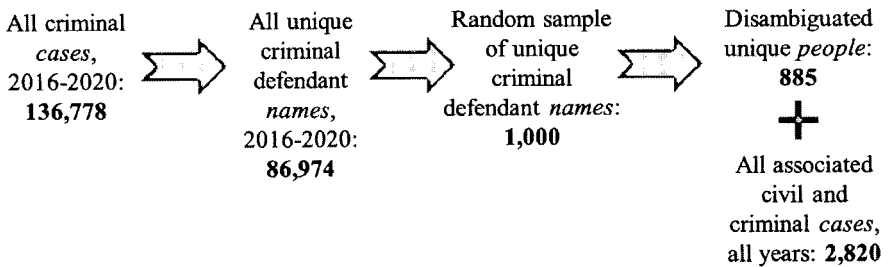
<sup>79</sup> For a discussion of the limits on the problem-solving approach in the family court context, see Jane M. Spinak, *Romancing the Court*, 46 FAM. CT. REV. 258, 266–69 (2008).

<sup>80</sup> In a call for courts to adopt common data standards that would govern organization of and access to court data, David Colarusso and Erika Rickard state, “Technologies aiming to disrupt the legal system in the service of access to justice should share the following principles: (1) collaboration between courts and other justice system partners; (2) user-centered design; and (3) openness to change existing practices when the need for and efficacy of change is supported by evidence.” David Colarusso & Erika J. Rickard, *Speaking the Same Language: Data Standards and Disruptive Technologies in the Administration of Justice*, 50 SUFFOLK L. REV. 387, 404 (2017).

ing on lessons from this process, in Part IV we offer proposals for restructuring and opening access to court data to better enable a people-first approach.

In broad strokes, our workflow began with *case-level data*,<sup>81</sup> pivoted to the *names* listed in the case data, took a random sample of those *names*, disambiguated the names to identify *unique people*, and then assembled all *cases* associated with each *unique person*. Put another way, we started with the courts' native data structure—using the case as the unit of analysis—and ended with a data set that is reorganized around the person. Figure 1 illustrates the process.

FIGURE 1: DATA ASSEMBLY WORKFLOW



### A. Pilot Study Site

Our pilot study draws data from three courts of original jurisdiction in Fulton County, Georgia: the Magistrate, State, and Superior Courts. The Magistrate Court handles civil matters in which the amount in controversy is less than \$15,000, as well as preliminary proceedings in criminal cases. The State Court handles all non-felony criminal cases and civil cases that do not otherwise fall into the Superior Court's exclusive jurisdiction.<sup>82</sup> The State Court also adjudicates traffic citations, which are classified as misdemeanor criminal offenses in Georgia.<sup>83</sup> The Superior Court, in turn, decides felony criminal cases and civil cases of all types.<sup>84</sup> It exercises exclusive jurisdic-

<sup>81</sup> For reasons explained further below, we began by assembling a list of criminal cases, which, unlike other case types, also include defendant demographics.

<sup>82</sup> See FULTON CNTY. STATE CT., <https://www.fultoncountyga.gov/inside-fulton-county/fulton-county-departments/state-court> [<https://perma.cc/E8QB-YGFE>] (last visited June 17, 2022) (listing civil matter types handled as including medical and legal malpractice, wrongful death, serious personal injury, product liability, and breach of contract cases and criminal matter types as misdemeanors, including simply battery, DUIs, criminal trespass, and traffic citations).

<sup>83</sup> See *id.*

<sup>84</sup> FULTON CNTY. SUPER. CT., <https://www.fultoncountyga.gov/inside-fulton-county/fulton-county-departments/superior-court> [<https://perma.cc/P3YR-8TY9>] (last visited June 17, 2022).

tion over real estate/land title and family and domestic relations cases.<sup>85</sup> The Superior Court also hears appeals from the Magistrate and State Courts.

Our study's geographic site, Fulton County, is Georgia's most populous county and the location of Atlanta, the state's capital and largest city.<sup>86</sup> Its courts are the busiest in Georgia and in the Southeast as a whole.<sup>87</sup> Despite their function as the gateway into the legal system for many litigants, county-level courts of original jurisdiction like Fulton's are greatly understudied even within courts-first scholarship in comparison with state appellate courts and federal courts.<sup>88</sup> By choosing a set of county-level courts as the site of our pilot study, we offer a small contribution toward filling that gap.

### B. Criminal Case Number Assembly

To assemble our data set, we used Fulton County's online court record search portal<sup>89</sup> to download a list of all criminal cases filed in all three courts in the period 2016–2020, or the previous five full years at the time we performed the work in summer 2021. We used criminal cases as our starting

<sup>85</sup> GA. SUPER. CTS., <https://georgiasuperiorcourts.org/question/what-is-the-difference-between-state-magistrate-probate-superior-courts/> [<https://perma.cc/UQ66-XMYF>] (last visited June 17, 2022).

<sup>86</sup> *Georgia Among Top Five Population Gainers Last Decade*, U.S. CENSUS BUREAU, (Aug. 25, 2021) <https://www.census.gov/library/stories/state-by-state/georgia-population-change-between-census-decade.html> [<https://perma.cc/YN5F-WEMF>] (listing Fulton County as the most populous county in Georgia); U.S. CENSUS BUREAU, *2020 Census Count by Georgia City Population*, (Aug. 12, 2021) [https://www.legis.ga.gov/api/document/docs/default-source/reapportionment-document-library/2020-census-count-by-city-population.pdf?sfvrsn=274928ba\\_2](https://www.legis.ga.gov/api/document/docs/default-source/reapportionment-document-library/2020-census-count-by-city-population.pdf?sfvrsn=274928ba_2) [<https://perma.cc/RYN9-Q7LW>] (listing Atlanta as most populous city in Georgia).

<sup>87</sup> *Court Services*, FULTON CNTY. CTS. SERVS., <https://www.fultoncountyga.gov/services/court-services> [<https://perma.cc/H8ZT-6DD2>] (last visited June 17, 2022).

<sup>88</sup> See, e.g., Carpenter et al., *supra* note 24; Alexander & Feizollahi, *supra* note 20 (discussing patchwork of state court electronic data access systems as a reason for their neglect in the courts-first scholarship). There is some limited precedent for this study in work by eviction researchers who have studied dispossession filings in Fulton County courts. See, e.g., ELORA RAYMOND ET AL., FED. RSRV. BANK OF ATLANTA, CORPORATE LANDLORDS, INSTITUTIONAL INVESTORS AND DISPLACEMENT: EVICTION RATES IN SINGLE FAMILY RENTALS (2016), <https://www.frbatlanta.org/-/media/documents/community-development/publications/discussion-papers/2016/04-corporate-landlords-institutional-investors-and-displacement-2016-12-21.pdf> [<https://perma.cc/L3LR-DBB7>].

<sup>89</sup> *Fulton County Magistrate, State, and Superior Court Record Search*, FULTON CNTY., GA., <https://publicrecordsaccess.fultoncountyga.gov/Portal/Home/Dashboard/29> [<https://perma.cc/R7RY-D5Y4>] (last visited June 17, 2022). Unlike some other counties, Fulton County has invested substantial resources in making court records publicly available via an online portal from which we could download records in bulk, contracting with popular court services vendor Tyler Technologies. See *Fulton County, Georgia, Selects Tyler Technologies to Provide Integrated E-Filing Services*, BUSINESSWIRE (Apr. 8, 2015), <https://www.businesswire.com/news/home/20150408005134/en/Fulton-County-Georgia-Selects-Tyler-Technologies-to-Provide-Integrated-E-Filing-Services> [<https://perma.cc/Z9ME-CECG>]. Since we originally assembled our data in summer 2021, the county made changes to the online portal that make extension or replication of our work more challenging. We discuss these changes further in Part IV below.

point because those case records include defendant demographic information (race, gender, and home city, state, and zip code) that civil case records lack.

Fulton County is not unique in this respect: individuals' demographic information is often absent from court data.<sup>90</sup> It can therefore be challenging to explore the individual litigant experience as informed by race or gender. When various court data sets are linked—as demonstrated by our pilot study methodology as well as the similar federal dyad-linking approach employed by other researchers and described in Part IV below<sup>91</sup>—demographic information available from one court or case type can be applied across the combined data set, enabling new forms of research. For these reasons, we chose criminal cases, and their associated defendant demographics, as the starting point for our data acquisition workflow.

We downloaded the 2016–2020 criminal case list by feeding all iterations of criminal case numbers into the courts' search portal and aggregating the results. Specifically, Fulton County criminal case numbers consist of the filing year, rendered as either two or four digits, followed by a two-letter code for the particular criminal case type (CP for Superior Court criminal, unindicted; SC for Superior Court criminal; CR for State Court, criminal), followed by a multi-digit case identifier. Using an asterisk as a search wildcard as shown in Figure 2 below, we searched for all possible combinations of years and criminal case type codes in our study period of 2016–2020. As of June 2022, however, the portal no longer allows the use of asterisk wildcard searches, meaning that our specific methods likely cannot be used for future research with this database.<sup>92</sup>

---

<sup>90</sup> See Rostain and O'Hara, *supra* note 22, at 496 (explaining why many civil court data lack information about the race, gender, or ethnicity of litigants).

<sup>91</sup> See *infra* Part IV.

<sup>92</sup> When a search using a wildcard asterisk was performed in June 2022, the following message appeared: "One or more fields was completed incorrectly: Wildcard searches using the (\*) character are not allowed for case number, case cross-reference number or citation number searches."

FIGURE 2: SAMPLE CRIMINAL CASE NUMBER SEARCH

Fulton County Magistrate, State, and Superior Court Record Search

Welcome, Charlotte

Smart Search \*Required

Search Criteria ?

\* Enter a Record Number or Name in Last, First Middle Suffix Format

16CR\*

Clear Submit

Advanced Filtering Options -

This process produced a results list like that shown in Figure 3 below, including case number, name, filing date, case type, and case status, which we copied and saved. We have redacted case numbers and names here to respect defendants’ dignity and preserve their privacy, for reasons we explore further in Part IV.B below.

FIGURE 3: SAMPLE CASE NUMBER SEARCH RESULTS LIST

Case Number	Name	Filing Date	Type	Status
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Dormant - Bench Warrant Issued
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed
[REDACTED]	[REDACTED]	01/05/2016	MISDEMEANOR	Case Closed

1 2 3 4 5 6 7 8 10 items per page 1 - 10 of 79 items

This process was partially automated, but with substantial researcher intervention. First, we frequently encountered an “I’m not a robot” pop-up window known as reCAPTCHA that requires the user to click a box to proceed. Our student assistants clicked through this box as part of our download pro-



cess.<sup>93</sup> Second, the number of results that a single case number search could return was capped at 200. We first ran our case number searches by month, and if monthly totals exceeded 200, we then ran the searches by day. For all except forty days, the per-day search results fell below 200. For those forty, we held constant the year and case type indicator as described above but varied the initial digits of the case identifier that followed rather than replacing them wholesale with a wildcard asterisk.

At the end of this workflow, we had assembled a data set consisting of 136,778 criminal cases filed in all three courts, 2016–2020. However, pivoting from a case-level view of court records to a person-level view required multiple additional steps. As an initial matter, the case records available via Fulton County’s portal contain no unique per-person identifier like a Social Security number or birth date (or even birth year)—and rightly so, as public access to such personal information could cause enormous harm to litigants. The courts could—but do not—assign their own publicly available unique identifier to individual litigants. The courts do appear to group at least some criminal cases under the same defendant name (see Figure 4 below). However, this linking does not extend to all case types.

As we argued in Section I.A, this lack of per-person identifiers or groupings illustrates that the courts themselves, in what data they make publicly available, do not adopt a people-first view of the judicial system. It also helps explain the dearth of people-first empirical research, as restructuring a per-case data set into a per-person data set becomes substantially more labor- and cost-intensive.

### C. Defendant Name Lookups

The next stage of our workflow pivoted from assembling a list of criminal *cases*—our starting point—to gathering the defendant *names* from those cases. With the names in hand, we could then assemble each *person’s* full civil and criminal case history, as well as demographic information derived from the criminal case records.

To accomplish this, we first harvested 86,974 unique defendant names from our criminal case list. We then took a random sample of 1,000 names. We proceeded with a sample rather than the full set because the reCAPTCHA challenge continued to appear when searching by name just as it did when searching by case number. This required manual intervention to

---

<sup>93</sup> We chose to hire, and fairly compensate, student research assistants to perform these tasks rather than engage the services of companies that offer to “defeat” reCAPTCHA and other similar devices. *See, e.g., Spammers Use The Human Touch To Avoid CAPTCHA*, NPR (Oct. 17, 2020), <https://www.npr.org/templates/story/story.php?storyId=130594039> [<https://perma.cc/PM8Q-GDXV>] (describing “CAPTCHA sweatshops” in which intermediary companies engage offshore workers at extremely low rates of pay to click through pop-up windows).

click the “I’m not a robot” box and prevented full automation of the data assembly process.<sup>94</sup>

Here, the 1,000 unique names do not necessarily represent 1,000 unique people: John Doe as a unique name with twenty associated criminal cases could be twenty different John Does with one case each or one John Doe with twenty cases. Likewise, John C. Doe and John Doe—two separate unique names appearing on a criminal case list—could be different people or the same person with different versions of their name entered in different criminal case records. We address this disambiguation problem at a later stage of our analysis and we describe our solution in the section that follows.

Proceeding with the 1,000 unique criminal defendant names, we wrote and deployed code to enter each name into the Fulton County courts’ search portal and gather each name’s full court case search results, which could include multiple criminal and civil cases of all types from all years, not confined to 2016–2020.<sup>95</sup> However, because our starting point was criminal cases filed between 2016 and 2020, all per-name results contained at least one criminal case filed in those years.

To illustrate, Figure 4 below shows a redacted sample from among the thirty-two results generated by searching for a single name in the format “Doe, Jane.”

---

<sup>94</sup> As of June 2022, the online portal now requires users to register in order to perform a search. Our testing suggests that the reCAPTCHA challenge does not appear on name lookups when the user is registered, suggesting that we could automate this portion of our methodology and increase the scope of this study in future iterations.

<sup>95</sup> We proceeded with a sample, rather than the full set of over 86,000 unique names, because the reCAPTCHA challenge continued to appear when searching by name just as it did when searching by case number. This required manual intervention to click the “I’m not a robot” box and prevented full automation of the data assembly process.

FIGURE 4: SAMPLE SEARCH RESULTS

XX/XX/XXXX					
Current Address: Ellenwood, GA 30294-2246		Race Black	Gender Female	Cases (2)	
Cases					
Case Number	Style / Defendant	File Date	Type	Status	
[REDACTED]	[REDACTED]	01/04/2016	NI-NON COMPLEX	Closed	
[REDACTED]	[REDACTED]	09/11/2015	NI-NON COMPLEX	Closed	
XX/XX/XXXX					
Current Address: COLLEGE PARK, GA 30349		Race Black	Gender Female	Cases (2)	
Cases					
Case Number	Style / Defendant	File Date	Type	Status	
[REDACTED]	[REDACTED]	09/24/2012	MISDEMEANOR	Case Closed	
[REDACTED]	[REDACTED]	08/06/2008	RN-PREFILE / CRIMINAL COMPLAINT	Open	
XX/XX/XXXX					
Current Address: Covington, GA 30016				Cases (1)	
Cases					
Case Number	Style / Defendant	File Date	Type	Status	
[REDACTED]	1ST FRANKLIN FINANCIAL VS. [REDACTED]	07/28/2014	CONTINUING GARNISHMENT	MAGISTRATE CASE CLOSED	

The first set of results shows criminal cases filed in 2015 and 2016 and, because this is a criminal case record, race, gender, and an Ellenwood, GA address. Notably, this result groups two cases under the same criminal defendant's name, providing some per-person linkage between cases. However, the second sample result for the exact same name lists criminal cases in 2008 and 2012 and an address in College Park—about a twenty-minute drive from Ellenwood. Did this defendant move from College Park to Ellenwood between 2012 and 2015? Or is this a different person?

The story is complicated further by the third result in Figure 4, a garnishment action, filed in 2014, with the exact same defendant name but a third city, Covington, much further to the east. The same questions arise: Are these one, two, or three separate people? Assuming away privacy problems, birth dates, which are listed at the upper right of both sets of criminal results as XX/XX/XXXX, could help resolve this for the criminal cases. However, the problem would remain for cases that lack birth dates or another identifier or grouping, such as the garnishment case shown in Figure 4.

Additional complexity comes from middle names and initials. For instance, the same first and last names as in the Jane Doe example appear in other case records with the following middle names or initials (altered for anonymity): D., L., S., Charmaine, Marie, Nicole, and Shannon. It is unclear whether "S." and "Shannon" should be assumed to be the same person, and

further, whether any of the above should be collapsed into any of the Jane Doe records that lack any middle name or initial.

Ultimately, our process yielded 3,252 civil and criminal case records of all types from all years associated with the random sample of 1,000 defendant names. Our next challenge was name disambiguation to attempt to solve the problems highlighted above and correctly cluster the names with each other and their associated cases.

#### D. Name Disambiguation

Broadly speaking, “disambiguation” refers to a process of developing rules applicable to sets of name variations to make a guess as to which groups of names—and here, their associated civil and criminal cases—refer to single unique individuals. This process necessarily introduces errors and privileges relatively uncommon or unusual names, both of which we discuss further below.

We experimented with multiple disambiguation approaches at different levels of complexity.<sup>96</sup> At the most complex, our method required the application of ten multipart rules and incorporated city and state, race and gender (when present), time elapsed between case filing dates, and a measure of how common or uncommon each first, middle, and last name was using census, Social Security, and other records from the United States, Canada, the United Kingdom, and Australia.<sup>97</sup> However, because this method relied heavily on administrative data from primarily English-speaking countries, it failed to account well for names that originate in non-English speaking regions of the world, and its reliance on demographic data limited its application only to criminal cases where such data were present.

We also experimented with a very simple approach, grouping together two sets of names:

- (1) Those in which first and last names were exactly the same and no middle initial or name appeared in any other variation of the name; and
- (2) Those in which first, middle, and last name were exactly the same, or in which middle name and middle initial matched, and no differ-

---

<sup>96</sup> Researchers have used a wide variety of other name and address disambiguation approaches, many of which leverage machine learning to group together similar, but not identical, names and addresses. *See, e.g.*, Georgetown University, *Civil Justice Data Commons, Maricopa County, AZ Justice Courts* (using Jaccard similarity and probabilistic soft matching to identify groups of eviction cases filed by the same landlord), <https://redivis.com/datasets/0q3m-f905yt0gd> [<https://perma.cc/MB58-MBRP>].

<sup>97</sup> *See Top 10 Baby Names of 2021*, SOC. SEC. ADMIN., <https://www.ssa.gov/OACT/babynames/> [<https://perma.cc/94VK-RPGB>]; Arun Rao, *Gender by Name Data Set*, U.C. IRVINE MACHINE LEARNING REPOSITORY <https://archive.ics.uci.edu/ml/datasets/Genderfy+Name> [<https://perma.cc/H65B-UU3N>] (combining name popularity records from the United States, United Kingdom, Canada, and Australia); *Frequently Occurring Surnames from the 2010 Census*, U.S. CENSUS BUREAU, [https://www.census.gov/topics/population/genealogy/data/2010\\_surnames.html](https://www.census.gov/topics/population/genealogy/data/2010_surnames.html) [<https://perma.cc/763D-BVY4>] (last visited June 17, 2022).

ent middle name, different middle name with the same first letter, or different middle initial appeared in any variation of the name.

If any variation appeared that fell outside of either of these two sets, the entire set of names was discarded from the data set. As an example, the Jane Doe example in Figure 4 above would fail this test because we could not be sure of the correct grouping of the names and their associated cases, given the presence of “Jane Doe” on its own, as well as “Jane Doe” with seven different middle names or initials. This approach was the most conservative, in that it did not group names and variations together unless our strict sequence of rules suggested substantial reason to do so.

We compared the output from all simple and complex rules-based groupings against manual groupings of a subset of names performed by our student research assistants, who used all available data as well as their own intuition about the names most likely to belong to the same person. Ultimately, we chose the simplest, and most conservative, rule-based disambiguation approach described above. This approach performed about as well as the more complex rule-based approaches and human review, and—given its conservatism—erred in favor of including only the most certain groupings in our final data set. After automating the application of our simple disambiguation approach, we performed one final layer of human review to identify and drop any duplicate case entries that were present in error on the Fulton County courts website and any name groupings that we still felt, via human judgment, were not reliably related to the same person.

We note that this process likely erred in favor of including less common names or name spellings and excluding common ones from our data set. For example, we omitted names such as the following from our data set, along with their associated civil and criminal cases: Robert Allen, David Johnson, Eddie Johnson, Mark Robinson, Gregory Taylor, and Stacy Wright. This perhaps introduced bias into our findings, as people with less common names or uncommon name spellings might differ systematically from those with more typical names and spellings.<sup>98</sup> If this project were expanded beyond the pilot described here, and if name disambiguation continues to be a key step in the workflow, researchers will need to estimate the size, direction, and implications of this bias. This is an additional point in support of our suggestion, explained in full in Part IV, that courts develop and make available reliable

---

<sup>98</sup> See, e.g., Yungu Kang, David H. Zhu, & Yan Anthea Zhang, *Being Extraordinary: How CEOs' Uncommon Names Explain Strategic Distinctiveness*, 42 STRAT. MANAGEMENT J. 462 (2020) (hypothesizing that “CEOs with uncommon names tend to develop a conception of being different from peers and accordingly pursue strategies that deviate from industry norms”); LaSonya L. Moore, Martha Lue Stewart, Dena D. Slanda, Anais Placencia, & MezNari M. Moore, *The Power of a Name: Nontraditional Names, Teacher Efficacy, and Expected Learning Outcomes*, 83 J. ENG. LEARNER EDUC. 83 (2020) (suggesting name difference as a possible harmful source of bias against nontraditionally-named children in classroom education).

unique identifiers or case linkages that would obviate the need for a name-dependent disambiguation workflow.

At the end of this step in our data assembly process, 885 unique individuals remained with 2,820 associated cases. Because each case grouping contains at least one criminal case—the starting point in the analysis—we were able to attribute the race, gender, and address data available in criminal case records to all cases in any given person-level grouping.<sup>99</sup> Beyond those fields, our data contain the following for each case: full case name, filing date, case type, case status, and role. As shown in Figure 4 above, “full case name” contains the opposing party’s name in civil cases; in criminal cases, only the defendant’s name appears. “Case type” in its original format contained fifty-seven values assigned by the various court clerks’ offices. We cleaned and combined the values into consolidated case types as shown in Appendix A for ease of analysis. At various points in the analysis, we dropped or included the “Appeal” and “Prefile” cases, depending on the question at hand. We note these choices in the sections below. In addition, case status values and distribution across case types are shown in Appendix B, but are not used in the present project, as the most common statuses listed were “closed” and “open,” which do not communicate the substantive outcome for the parties involved. Later work may incorporate per-person case outcome data into the analysis if it can be derived from other sources like case docket sheets.

Finally, we assigned a role to each person in each civil and criminal case: plaintiff, defendant, and child. We derived the “plaintiff” and “defendant” roles from the name position before or after the “versus” in the full case name in civil cases or, for criminal cases, the single name listed in the full case name field. The “child” role is applicable to child support cases brought by the Georgia Department of Human Services, where a parent is named in the full case name and the child beneficiary is listed separately, and the child—presumably as an adult—later appears in other types of civil or criminal case.

Before proceeding to descriptive statistics and findings, a word on what our data do not contain. Our methodology excludes people whose entire set of touchpoints was civil, and those who had criminal cases exclusively outside the 2016–2020 window. In addition, the data set we produced was generated from a random sample of criminal defendant names, so is not comprehensive even as to our scope. Finally, we state the obvious point that our data do not reach courts not included in Fulton County’s online portal, but where some of the people in our people-first approach might have proceedings. This includes the Fulton County Probate Court, which decides

---

<sup>99</sup> Four people in our data set had a single civil case only. Though these four originally entered the data set with a criminal case as well, we dropped those criminal case filings as erroneous, as they were either duplicates or misattributed to the person at issue. The four remain in our total of 885 people but, because they have single cases only, do not figure into our case sequence analyses presented in Part III.B below.

cases involving wills, estates, guardianships, and conservatorships,<sup>100</sup> as well as courts elsewhere in the Georgia state and federal systems. As noted above, the fractured and siloed nature of the U.S. legal system—and its myriad court recordkeeping systems—renders a truly comprehensive people-first view of court data extraordinarily difficult to obtain. Despite these limitations, our pilot study offers a useful road map for how people-first data might be generated and used by researchers and how reform-minded court administrators might restructure the data they produce to enable further people-first analysis.

### III. FINDINGS

This Part first presents a set of descriptive statistics on our pilot study data, using a people-first approach as our means and manner of reporting. We then offer some findings relevant to the four research areas described above: collateral consequences, holistic legal services, problem-solving courts, and eviction.

#### A. *What Do “People-First” Data Look Like?*

Our data set comprises 885 people and 2,820 associated civil and criminal cases, with at least one criminal case per person filed in the five-year period between 2016-2020. The average number of cases per person is 3.2. Forty-six percent of people had only one case; the remainder had anywhere between two and twenty-eight, filed as early as 1971 and as late as 2021. Figure 5 shows the distribution of number of cases over the sample of people. Figure 6 shows the time elapsed from first to last case.

---

<sup>100</sup> See *Probate Ct. of Fulton County*, FULTON COUNTY <https://www.fultoncountyga.gov/inside-fulton-county/fulton-county-departments/probate-court> [<https://perma.cc/KK7J-KLN5>].

FIGURE 5: HISTOGRAM OF NUMBER OF CASES PER PERSON

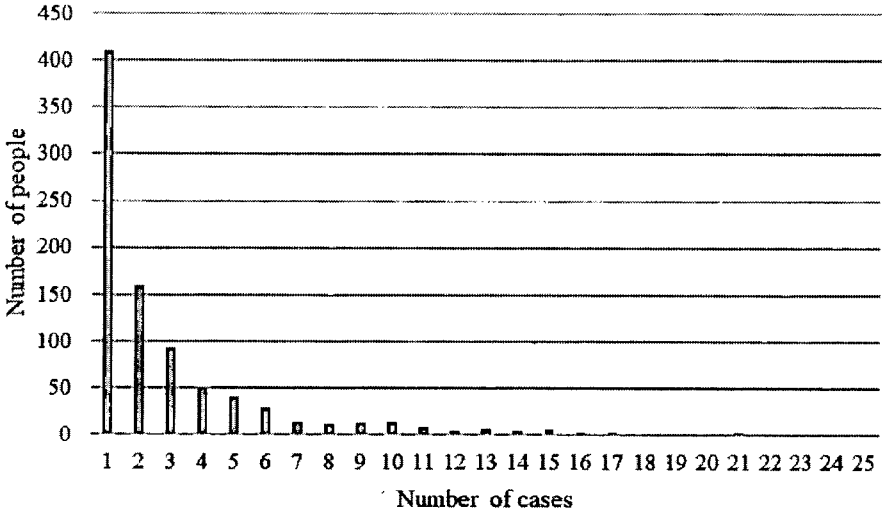
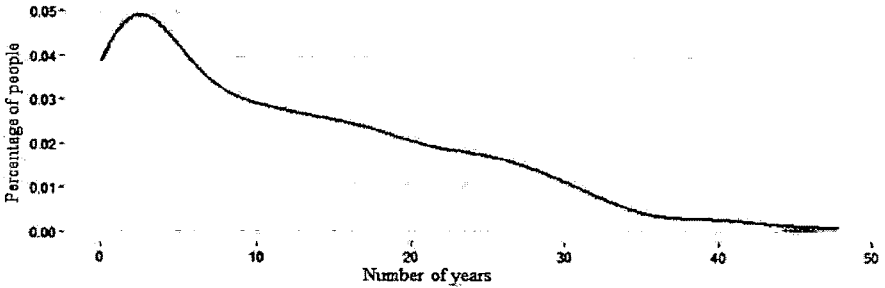


FIGURE 6: TIME ELAPSED BETWEEN FIRST AND LAST CASES PER PERSON



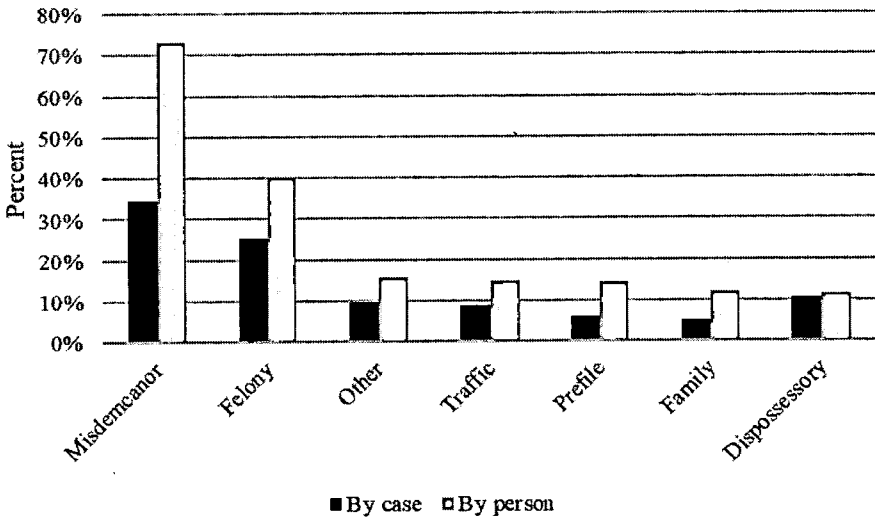
As Figures 5 and 6 suggest, our data set is roughly split between people with single cases and those with multiple, and multiple-case sequences mostly clustered in a time period of fewer than five years. The outliers—those people with many more cases in sequence or cases spread over a longer period of time—may be different in kind as well as in quantity from their single-case counterparts. We explore these differences further in the sections that follow but are hampered by access to only limited person-level demographic data. Nevertheless, this preliminary observation alone—that there may be meaningful subgroups of people who have contact with the courts, demarcated by their number and concentration of court touchpoints—is the type of policy-relevant finding that a people-first data approach enables.

Within our data set, misdemeanors were the most frequent case type, representing thirty-five percent of all cases and seventy-three percent of all



people, followed by felonies at twenty-five percent of all cases and forty percent of all people.<sup>101</sup> Figure 7 shows the full distribution of cases, shown by the darker set of bars, and people, shown in the lighter set, across the major case type categories. “Dispossessory,” shown in the right-most bar in Figure 7, is the term in Fulton County for an eviction case. Unsurprisingly, criminal case types dominate: recall that our data do not capture all people with cases in Fulton County’s courts, but only those who had at least one criminal case of any type<sup>102</sup> filed in the 2016–2020 period.

FIGURE 7: DISTRIBUTION OF CASES AND PEOPLE BY CASE TYPE



Continuing with a people-first framework, of the 885 people in our data set, the largest proportion (74.1%) was Black or African-American, followed by white (19.9%), followed by other races and ethnicities at a substantial remove. Table 1 below lists all race/ethnicity categories captured by the courts—using the courts’ terminology and based on litigant self-reporting during the intake process<sup>103</sup>—and their representation within our pool of 885

<sup>101</sup> Here, individuals were often associated with more than one of any given case type, explaining the per case and per person differentials.

<sup>102</sup> Criminal case types include felony, misdemeanor, prefile, and traffic. The “prefile” category captures criminal complaints that appear not to have continued into full-fledged criminal prosecutions.

<sup>103</sup> According to a court clerk interviewed by one of our research assistants, race data in the courts’ criminal records systems are pulled from jail records; arrestees self-report their race during intake at the Fulton County Jail. *Interview with T.B., Judicial Assistant, FULTON COUNTY COURT* (Oct. 31, 2022); see also *Pretrial Intake Overview*, SUPERIOR COURT OF FULTON COUNTY, <https://www.fultoncourt.org/intake-unit> [<https://perma.cc/5AQC-XDW6>] (“The Pretrial Services Intake Unit collects, verifies, and reports to the Court Officer background information and criminal histories on individuals charged with felony and or misde-

litigants. The table also reports the percentage of all Fulton County residents by race and ethnicity according to 2021 population estimates by the U.S. Census Bureau.<sup>104</sup> This is necessarily an inexact comparison but provides some benchmarking. As the comparison reveals, Black people are over-represented among the sample set of 885 litigants in our data set, while every other population group for which there is data available was under-represented as compared to the county population as a whole. This is a matter of concern but perhaps unsurprising given the predominance of criminal cases in our data set and the racial disparities that have been well-documented in every part of the U.S. criminal legal system.<sup>105</sup>

TABLE 1: RACE AND ETHNICITY OF LITIGANTS VERSUS FULTON COUNTY RESIDENTS

<i>Race/Ethnicity</i>	<i>Frequency</i>	<i>Percent</i>	<i>Census Estimate</i>
Black	656	74.1%	44.7%
White	176	19.9%	44.9%
Unknown	32	3.6%	-
Asian	8	0.9%	7.8%
Other	7	0.8%	-
Native Hawaiian or Other Pacific Islander	2	0.2%	Z <sup>106</sup>
Hispanic	1	0.1%	7.3%
Indian	1	0.1%	0.3%
Middle Eastern	1	0.1%	-
Multiracial	1	0.1%	2.3%

Layering in case type data, we can further investigate racial and ethnic disparities. Table 2 reports the percentage of each case type, shown at a more granular level, associated with litigants by race and ethnicity, sorted by the differential between Black and white litigants, the two largest racial groups represented in the data set. While large racial differentials in criminal case

meanor offenses. This information is gathered on eligible defendants booked into the Fulton County jail before their First Appearance Hearing.”).

<sup>104</sup> *QuickFacts: Fulton County, Georgia*, U.S. CENSUS BUREAU, <https://www.census.gov/quickfacts/fultoncountygeorgia> [<https://perma.cc/5WNH-LGWS>] (last visited Aug. 8, 2022) [hereinafter “QuickFacts”].

<sup>105</sup> *Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System*, THE SENTENCING PROJECT (Apr. 19, 2018), <https://www.sentencingproject.org/publications/un-report-on-racial-disparities/> [<https://perma.cc/5BFP-RMZE>].

<sup>106</sup> The census data list the percentage of Native Hawaiian or Other Pacific Islander as “Z,” which denotes “Value greater than zero but less than half unit of measure shown.” In other words, the percentage would be less than 0.5 percent. See QuickFacts, *supra* note 104.

types shown in Table 2 track the nationwide trends discussed above, the family, dispossessory, and traffic case types also display large Black-white differentials. This could be an artifact of the data acquisition process, which began with criminal cases—a category in which Black people are over-represented—as its starting point, meaning that every case type is likely also characterized by Black overrepresentation. But switching to a people-first lens, as we do in the sections that follow, has the potential to reveal other insights about the sequencing and clustering of case types, and the potential collateral consequences of criminal case involvement in other areas of law.

TABLE 2: RACE AND ETHNICITY OF LITIGANTS BY GRANULAR CASE TYPE

<i>Case Type</i>	<i>Black</i>	<i>White</i>	<i>Unknown</i>	<i>Remainder</i>	<i>Black-White Differential</i>
Prefile	90.6%	7.6%	0.0%	1.8%	82.9%
Small Claims	90.9%	9.1%	0.0%	0.0%	81.8%
Felony	89.1%	8.8%	0.3%	1.8%	80.2%
Family	87.1%	7.5%	5.4%	0.0%	79.6%
Tort	84.2%	10.5%	2.6%	2.6%	73.7%
Abandoned Motor Vehicle	84.1%	12.7%	3.2%	0.0%	71.4%
Other	78.3%	19.6%	2.2%	0.0%	58.7%
Dispossessory	77.6%	19.7%	1.7%	1.0%	58.0%
Misdemeanor	76.0%	19.0%	2.6%	2.5%	57.1%
Traffic	72.1%	25.1%	0.0%	2.8%	47.0%
Contract	65.6%	31.3%	1.6%	1.6%	34.4%
Garnishment	53.6%	46.4%	0.0%	0.0%	7.1% <sup>107</sup>

The maps in Figure 8 below reveal yet another worrisome, though perhaps again unsurprising, pattern. While the cases in our data set are concentrated in the southern half of Fulton County, as shown by the left panel, high incomes are concentrated in the north, as shown by the panel on the right.<sup>108</sup> This may suggest, at minimum, a positive correlation between low income and court involvement, particularly involvement with the criminal legal system—another relationship that has been well-established in the literature.<sup>109</sup>

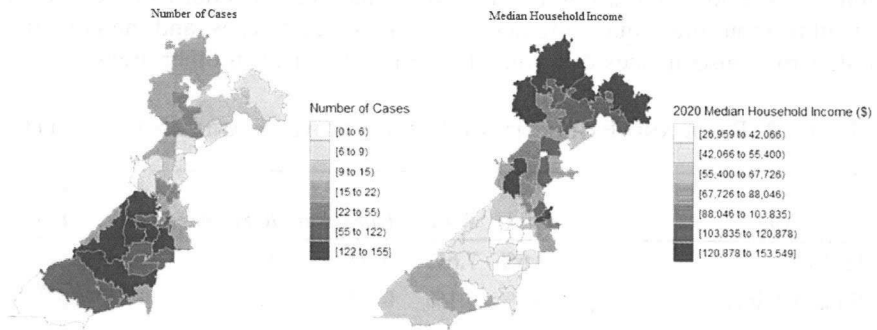
<sup>107</sup> Interestingly, the racial differential is the smallest for the “garnishment” case type, or a legal proceeding for recovery of a debt from a judgment debtor. *See Garnishments*, MAGISTRATE COURT, FULTON COUNTY, GA, <https://www.magistratefulton.org/183/Garnishments> [<https://perma.cc/2Z6E-SRFJ>]. Further research into the nature of these cases—and the characteristics of the people involved—may shed interesting light on this finding.

<sup>108</sup> A more nuanced analysis would adjust case filing numbers per capita and would account for all cases filed of all types rather than the small pilot set used here.

<sup>109</sup> Researchers consider the causal arrow as pointing in both directions between poverty and involvement in the criminal legal system. *See, e.g.,* Terry-Ann Craigie, Ames Grawert,

In our pilot study, we did not have access to individual-level income information, so we cannot perform a true person-level analysis of income-based correlations. Nevertheless, this case-level view is powerfully suggestive as a starting point for further person-level examination.<sup>110</sup>

FIGURE 8: CASES BY ZIP CODE AND MEDIAN HOUSEHOLD INCOME<sup>111</sup>



With this person- and case-level orientation as background, we now turn to our findings relevant to the four substantive research areas introduced above: collateral consequences, holistic legal services, problem-solving courts, and eviction.

### B. Research Impact of People-First Data

In broad terms, each of the four substantive research areas is concerned with causation, or the idea that involvement with the courts may trigger or be triggered by other court involvement or by factors outside the courthouse walls. The collateral consequences stream of research attempts to identify and study those chains of events, beginning with either a criminal or civil

---

Cameron Kimble, & Joseph E. Stiglitz, *Conviction, Imprisonment, and Lost Earnings: How Involvement with the Criminal Justice System Deepens Inequality*, BRENNAN CTR. FOR JUST. (Sept. 15, 2020), <https://www.brennancenter.org/our-work/research-reports/conviction-imprisonment-and-lost-earnings-how-involvement-criminal> [<https://perma.cc/SW4D-CLRE>] (finding positive relationship between lifetime lost earning capacity and encounters with the criminal legal system resulting in conviction and imprisonment; also exploring poverty as a predictor of criminal legal system involvement).

<sup>110</sup> We performed similar analyses on the basis of sex. The court data recorded “male” as the sex of just over 71.4% of people in the data set. Men outnumbered women in all of the case types. Only “small claims” was close, with men accounting for fifty-five percent of cases and women forty-five percent. This distribution is interesting, suggesting an opening for future people-first research into both sexes’ use of small claims court as a dispute resolution forum.

<sup>111</sup> The geographical units in the maps in Figure 8 are Zip Code Tabulation Areas, or ZTCAs. Median income data comes from the 2020 U.S. Census. *Median Income by ZTCA*, U.S. CENSUS BUREAU, [https://data.census.gov/cedsci/table?q=median%20income%20by%20zcta&tid=ACSST5Y\\_2020.S1901](https://data.census.gov/cedsci/table?q=median%20income%20by%20zcta&tid=ACSST5Y_2020.S1901) [<https://perma.cc/F5LN-HEHN>] (searching with income and zip code filters applied). ZCTAs 30336, 30334, and 30332 were missing data for 2020; the map above replaces 2020 data with 2019.

case as the first link. Holistic legal approaches focus on the whole client, who may experience an array of interrelated sequential or simultaneous legal and nonlegal problems. Advocates attuned to those relationships, the thinking goes, provide more effective client representation if they address the contingent and complex nature of legal problems either directly or by recruiting additional support services and resources. Problem-solving courts, too, are built on an assumption of causality, attempting to understand the reasons for recidivism and solve the underlying problems that may lead parties to return to court. Finally, many scholars studying eviction focus on eviction's consequences, either in the form of additional legal proceedings or other types of harm to tenants and their families. In this way, the eviction research stream is functionally an inquiry into eviction's collateral consequences.

As summarized in Part I above and explored briefly in the previous section, courts-first data organized solely around the case as the unit of analysis do little to shed light on the causal chains that might link a particular person's or group of people's criminal cases to subsequent family or eviction cases, for example. All that a researcher could tally would be the numbers of criminal, family, and dispossession matters filed in a given court or courts, without observing any person-level linkages among and across case type categories. Restructuring the data, as we do here, to center the person enables exploration of person-level case sequences with an eye toward causation—the foundational assumption and core topic of inquiry of the four research areas covered here.

Indeed, as Figures 5 and 6 show above, the overall picture painted by our pilot study is one of repetition and return: over half of the 885 people in the study were involved in more than one case, and most people's cases occurred within a relatively compressed timeframe. While our findings below only demonstrate correlation, these compressed timeframes may be suggestive of the types of causal relationships between and among cases posited by researchers. Widespread availability of people-first court data would enable further study in this vein.

### *1. Collateral Consequences*

As noted above, scholars have applied the collateral consequences framework to both civil and criminal case chains; we do the same here. However, for manageability, we limit our analysis to people with case sequences that include at least one felony or misdemeanor—the two most common case types in our data set. We then track the civil and criminal cases that both precede and follow felonies and misdemeanors.

We first isolated all people with a felony or misdemeanor and at least one additional immediately neighboring case, or 50.7% of the 885 people in the data set. Overall, only twenty-nine percent of felonies were single cases, as were fourteen percent of misdemeanors, meaning that large majorities of

people with either case type also had other types of court involvement. Because our study began with criminal case filings in the 2016–2020 period, we cannot compare these single-case percentages to other case types, as our pilot study design did not include people with single civil cases. Investigating the single-case percentages across case types, however, would be a simple extension of our work, if person-first court data were more widely available.

Returning to our workflow, we focused first on immediate case pairs that began with either a felony or misdemeanor, regardless of where that case pair appeared within a person’s entire case sequence. Anchoring the analysis on criminal cases as potential triggers of future legal system involvement aligns with the extensive research summarized in Part I.B.2 above on the collateral consequences of criminal convictions. Thus, in this analysis, a person whose case history consists of felony-family would have one case pair in our felony analysis; a person with dispossession-felony-traffic-felony-felony would have two felony case pairs. The arc diagrams in Figures 9 and 10 below illustrate the second cases appearing in all felony and misdemeanor case pairs, respectively. The weight of the arced lines represents the frequency of the case pairs, while the horizontal axis on each diagram records the median time elapsed, in years, between the first and second case in the pairs.

FIGURE 9: FELONY CASE PAIRS WITH ELAPSED TIME

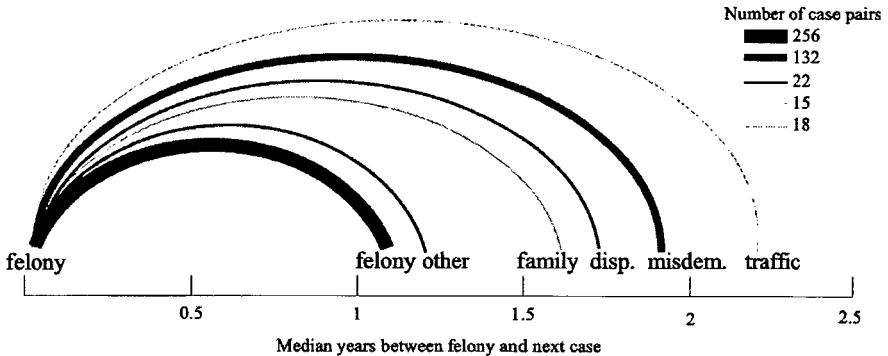
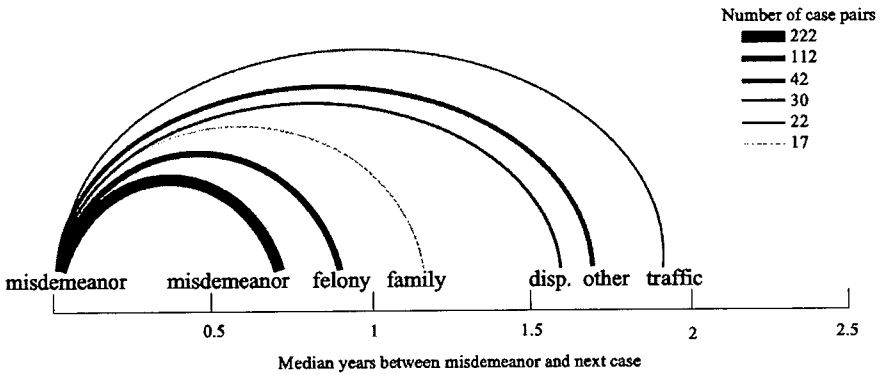


FIGURE 10: MISDEMEANOR CASE PAIRS WITH ELAPSED TIME



As the diagrams show, contingent on a person’s having multiple cases in sequence, a felony as a first case in a pair is most likely to be followed by another felony, just over a year later. The year timeframe is important, because in Georgia a felony is defined as a crime punishable by at least one year in prison.<sup>112</sup> One interpretation of these results is that defendants convicted of felonies are reoffending almost immediately after serving their sentences. However, recall that our data are based on case filings and do not reliably record dispositions, sentences, and time served; they do not indicate whether the felony case in fact resulted in a conviction or sentence. A more accurate interpretation of these results, therefore, is that felony *charges* cluster with other, quickly occurring felony *charges*. Regardless, this pattern is highly relevant to the concerns that motivate holistic legal services and problem-solving courts, suggesting a need for further study of felony-felony case pairings with access to more detailed person- and case-level information.

Figure 10, in turn, reveals similar patterns, showing misdemeanor-misdemeanor and misdemeanor-felony pairings as both the most frequent and the closest-occurring pairs in time. A larger person-level study could investigate whether and the circumstances under which misdemeanor case filings act as gateways to more serious criminal involvement in the form of felony filings, versus the misdemeanors that remain as part of less serious—though perhaps no less disruptive to family and economic stability—chains of subsequent misdemeanor filings. Here, as suggested above, researchers could test whether identifiable subgroups of people exist within the data, characterized by different patterns of case sequences, and suggesting different types of support and diversion interventions.

This person-level case pairing analysis also enables investigation of the civil case types that *precede* subsequent felony or misdemeanor case filings. Though causal claims are beyond the scope of this pilot study, knowing

<sup>112</sup> GA. CODE ANN. § 16-1-3(5) (2018) (“‘Felony’ means a crime punishable by death, by imprisonment for life, or by imprisonment for more than 12 months.”).

about common case patterns may aid court workers, attorneys, and other service providers in crafting supports and interventions to avoid escalation from civil to criminal court involvement. Here, we weighted the frequency of all case pairs in our data set by the median time lapse between the cases in the pairing, producing a ranking that combines the frequency and time lapse measures.<sup>113</sup> Table 3 below lists the top ten case pairs that involved a felony or misdemeanor in either the first or second spot, ranked by the combined frequency-median time lapse metric.

TABLE 3: TOP TEN CASE PAIRS INVOLVING FELONY OR MISDEMEANOR, RANKED BY COMBINED FREQUENCY-MEDIAN TIME LAPSE MEASURE

<i>Case Pair</i>	<i>Frequency</i>	<i>Median Time Elapsed (in years)</i>
Felony-felony	256	1.1
Family-felony	40	6.5
Felony-misdemeanor	132	1.9
Traffic-misdemeanor	76	3.2
Other-misdemeanor	64	2.4
Misdemeanor-misdemeanor	222	0.7
Family-misdemeanor	33	3.9
Misdemeanor-felony	112	0.9
Dispossessory-misdemeanor	46	2.1
Traffic-felony	26	3.5

Looking beyond the felony-misdemeanor pairings already explored in connection with Figures 9 and 10 above, the top ten in Table 3 include pairings between family law and both criminal case types (though the family-felony pair features a relatively long time lapse period of over six years) as well as dispossessory-misdemeanor. As above, both case sequences might signal possible early intervention points to prevent a transition from civil to criminal court involvement.

Traffic cases, too, appear among the top ten as felony or misdemeanor predecessors. Here, interpretation is difficult, as traffic violations may be so widespread as to be meaningless or may be signals of greater future involvement with the legal system. Indeed, a U.S. Supreme Court decision grants police officers the authority to use pretextual or technically permissible traf-

<sup>113</sup> Specifically, we indexed both measures to a common starting point and then multiplied the results together.



fic stops to pursue suspicions of other criminal activity.<sup>114</sup> Given the limited data in our pilot study, we are hesitant to suggest a single interpretation of these traffic-related findings, but highlight them as an opportunity for further study if people-first court data were to become more widely and easily accessible.

## 2. *Holistic Legal Services and Problem-Solving Courts*

The preceding sections have already suggested multiple ways in which people-first data could support the wrap-around client representation and advocacy that holistic legal services programs employ, as well as the whole-person approach that animates problem-solving, specialized, and unified courts. Indeed, the very concept of civil and criminal collateral consequences, with possible causal arrows pointing in many directions within a person's sequence of cases, suggests a role for both holistic legal services and problem-solving courts.

In this section, we offer one additional finding derived from our person-centered approach to illustrate the potential of people-first data to advance the holistic legal services and problem-solving projects. We introduced this finding above in connection with Haney's research on the connection between fathers' incarceration and child support actions; we return to it here.

There are seventy-three people in our data set with at least one child support case, for a total of ninety-seven child support actions. In some Fulton County court records, the child support defendant is listed along with the name of the child on whose behalf the support is sought. In thirty-eight child support cases, or thirty-nine percent of the ninety-seven total child support actions, the listed child also appeared in the data set as a litigant in at least one later case, in a total of 194 subsequent cases. We acknowledge the possibility of data entry errors here, as the Fulton County courts may have swapped the child support defendant's (usually father's) name with the child's name at various points, or a father and son might share the same name. These caveats point to the hazards of people-first research—like the present pilot study—conducted based on name matching. They also provide further support for our proposal that courts assign unique identifiers to the people who interact with the legal system. We discuss this proposal further

---

<sup>114</sup> See generally *Whren v. United States*, 517 U.S. 806 (1996). Researchers have also documented racial disparities in officers' exercise of discretion in initiating a traffic stop. See, e.g., Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff, & Sharad Goel, *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 *NATURE HUM. BEHAV.* 736, 736 (2020) (finding in a study of records from nearly 100 million police traffic stops nationwide that Black drivers were more likely to be stopped by police during daylight than at night, when their race was apparent; finding that Black and Hispanic drivers were more likely to be subject to car searches after stops than White drivers). This may explain, at least in part, the high Black-white differential shown in the "traffic" case type in Table 2 above.

in Part IV. Nevertheless, Table 4 below details the distribution of the subsequent cases associated with the children listed in child support cases, with the two main criminal case types ranking at the top along with dispossessionary actions.

TABLE 4: DISTRIBUTION OF SUBSEQUENT CASE TYPES FOR CHILDREN LISTED IN CHILD SUPPORT CASES

<i>Case Type</i>	<i>Percent of Subsequent Cases</i>
Felony	27%
Dispossessionary	26%
Misdemeanor	23%
Other	17%
Family	7%

While the figures in Table 4 come from small numbers and are far from definitive in establishing causation, they suggest a possible narrative: that children whose parents or guardians are involved in a child support dispute may be at risk themselves for later criminal legal involvement and/or economic instability, as suggested by the relatively high felony, misdemeanor, and dispossessionary numbers.

Holistic legal services providers might incorporate these types of people-first observations in their representation and advocacy for entire families involved in child support disputes, to mitigate the potential collateral consequences for both children and adults. Further, problem-solving, specialized, and unified courts, particularly those that have been set up as family law courts,<sup>115</sup> could use such data to guide their approach to adjudicating child support disputes, attending more closely to the child's or children's continued stability as they grow older and risk other types of court involvement. In that sense, a people-first approach might take on an even broader interpretation—reaching not only all of the legal issues affecting one person, but the other people affected as well.

### 3. *Eviction*

Our final set of people-first findings pertain to the literature on eviction filing patterns, and particularly other researchers' observations about the use and impacts of serial eviction filings as a rent collection device for landlords.<sup>116</sup> This work is a variation on the same causation-related (or correlational) themes explored thus far: the concern is with the downstream or

<sup>115</sup> See Ross, *supra* note 67, at 1 (discussing specialized family courts).

<sup>116</sup> See Leung et al., *supra* note 69, at 1 (studying the impact of serial eviction filings).

collateral effects of eviction or dispossession filings and the underlying instability that such filings suggest. We explore the possibility of subsequent impacts using our person-level case sequence data.

Twelve percent of people in our data set had at least one dispossession case filing, and dispossessions accounted for ten percent of all cases. (We suspect that, as reiterated elsewhere, this number would be higher if not for the limiting premise that the study only includes people with one or more criminal case.) Further, as Table 5 shows, the dispossession case type was associated with the highest mean (average) number of cases per person of all case types. This high person-level case volume may be a signal of serial eviction filings.

TABLE 5: MEAN NUMBER OF CASES PER PERSON, BY CASE TYPE

<i>Case Type</i>	<i>Number of cases</i>	<i>Number of people</i>	<i>Mean cases per person</i>
Dispossession	295	103	2.86
Felony	713	354	2.01
Other	272	138	1.97
Traffic	247	130	1.90
Misdemeanor	976	646	1.51
Family	147	106	1.39
Prefile	170	128	1.33

To investigate varying case filing patterns within our set of dispossessions, we created four mutually exclusive classifications:

- Single: A single dispossession case per person;
- Serial: A series of dispossession cases filed by the same landlord against the same tenant;
- Serial plus: A set of serial dispossession cases filed by the same landlord against the same tenant plus at least one additional dispossession filed by a different landlord;
- Multiple: Multiple dispossession cases filed by more than one landlord against the same tenant, with no set of serial dispossessions.

Table 6 below reports the percent of all dispossessions captured by each of these four categories, using a case-based framework, and the proportion of all people with at least one instance of each type of dispossession, using a person-level frame.

TABLE 6: DISPOSSESSORY TYPES AS A PERCENT OF DISPOSSESSORY CASES AND PEOPLE

<i>Dispossessionary Type</i>	<i>Percent of Cases</i>	<i>Percent of People</i>
Single	19%	50%
Serial (same landlord)	19%	16%
Serial plus	53%	24%
Multiple (different landlords)	10%	10%

As Table 6 shows, half of the people in our data set who had been a defendant in any dispossessionary case at all had been subject to more than one filing. A full forty percent (summing the “serial” and “serial plus” categories) experienced what researchers would classify as serial dispossessionaries, or repeated eviction filings by the same landlord against the same tenant. People who fell into these two groups experienced an average of 3.1 and 6 eviction filings in the serial and serial plus categories, respectively. An examination of the time lapse between dispossessionary cases adds to this story: the mean time elapsed from the first to the last case in the serial sequences was 178 days, or about two-thirds of a year, suggesting that landlords may have been engaging in concentrated, iterative dispossessionary filings but no actual evictions.

Turning now to eviction filings’ collateral consequences, we explored the case types that precede and follow all four categories of dispossessionary. Here, we note that the lack of reliable case outcome data prevents us from distinguishing between dispossessionary filings that resulted in an actual loss of housing and those that did not, such as the rapid-fire case filings examined above. A more comprehensive set of person-first data would include substantive case outcome information beyond the uninformative “open” and “closed” labels that dominate the present data set,<sup>117</sup> enabling researchers to pull apart and investigate the separate effects of the filing itself versus the filing’s on-the-ground outcome.

A second note of caution stems from our data assembly methods, which began with criminal cases filed in 2016–2020. Therefore, the dispossessionary findings relayed in this section do not capture all dispossessionaries filed in all years in Fulton County courts, but rather those filed against tenants who also themselves had a criminal case in the relevant five-year time period.

With these caveats in mind, we performed a case pairing analysis similar to the one described above in connection with felonies’ and misdemeanors’ collateral consequences. Table 7 shows the percent of each dispossessionary type that was followed, within one year, by any case of any other type, by any criminal case, by felony cases specifically, and by family

<sup>117</sup> See Appendix B.

cases. In other words, this analysis treats each instance of each category of dispossession as its own single occurrence, so serial, serial plus, and multiple dispossessions each become a single first instance in a case pairing analysis.

TABLE 7: PERCENT OF DISPOSSESSORY TYPE FOLLOWED BY SUBSEQUENT CASE WITHIN ONE YEAR, BY SUBSEQUENT CASE TYPE

<i>Dispossession Type</i>	<i>Followed by:</i>			
	<i>Any Case</i>	<i>Any Criminal</i>	<i>Felony</i>	<i>Family</i>
Single	30%	20%	8%	4%
Serial (same landlord)	56%	13%	6%	13%
Serial plus	58%	8%	4%	13%
Multiple (different landlords)	80%	50%	20%	0%

As Table 7 shows, sequences of multiple filings by different landlords are the most likely to be followed by any case of any other type, as well as by criminal cases and felonies in particular. Both serial dispossession types are the most likely to be followed by family law cases. Without further person-level data, and due to the small size of our pilot study, these results are difficult to interpret. Nevertheless, they may suggest that while serial dispossession filings by the same landlord are a destabilizing force in tenants' lives—suggested by subsequent family law cases—the presence of a sustained set of eviction filings by multiple different landlords suggests even greater disruption. These tenants are moving repeatedly, and this instability may result in criminal case involvement, or the moves and the criminal cases may both be symptoms of a common underlying problem. Regardless, people-first data could enable both holistic legal services providers and specialized housing courts to spot multiple-dispossession patterns and bring greater resources to bear to prevent escalating involvement with the legal system. In addition, there are likely many valuable insights regarding social and financial dynamics that researchers from other fields—including economics, sociology, public health, political science, and urban planning—could glean from such data, or gain by complementing their own work with such data.<sup>118</sup>

<sup>118</sup> For example, housing scholars might use such data to help explain the challenges that individuals interacting with any aspect of the legal system have in obtaining public or affordable housing, and how those dynamics correlate with data regarding eviction, or the loss of such housing. See, e.g., Peter Leasure, R. Caleb Doyle, Hunter M. Boehme, & Gary Zhang, *Criminal History, Race, and Housing Type: An Experimental Audit of Housing Outcomes*, 49 CRIM. JUST. & BEHAV. 1536, 1537 (2022) (studying “the impact of various types of multiple conviction records on private housing outcomes”); Peter Leasure, *Securing Private Housing With a Criminal Record*, 58 J. OFFENDER REHAB. 30 (2019) (studying interactions between tenants' past court interactions and future housing stability). Scholars might also study the impact on voting patterns or disease transmission caused by tenants' multiple moves due to evictions. See

Shifting the lens now to the time *before* a dispossession is filed, we found that felonies and misdemeanors were the most common case types appearing immediately prior to eviction filings, followed by traffic and family. These findings are also relevant to the substantive research areas discussed above: assuming causation, an eviction filing or eviction itself becomes a collateral consequence of criminal case involvement; holistic legal services providers could be attuned to the risk of eviction after criminal case filing; problem-solving courts could account for the potential for criminal case involvement to increase instability and trigger a dispossession, which could then make reoffending and recidivism more likely.

\*\*\*\*\*

This Part has demonstrated the potential for people-first court data to contribute to a wide variety of research and policy areas. However, this Part is also replete with warnings, cautions, and caveats stemming from the incomplete nature of the data available for our pilot study; the uncertainties introduced by name-based person and case matching in the absence of unique per-person identifiers; and the study's small size, necessitated by the resources required for the data assembly process. The next Part turns to our suggestions for reform to court data organization and access, which would address these problems and allow courts to become key partners in a people-first research and policy agenda.

#### IV. RESTRUCTURING COURT DATA

This Part provides recommendations as to how courts might restructure their data so that the people who work within them, the litigants who are subject to their authority, and the researchers who study them can more easily access data that speak to the individual experience. Subsequently, it addresses a number of concerns about people-first data—including diminished privacy, dignitary harms, and the possibility that such data might be used to target individuals based on their interactions with the court. While significant, these concerns are not insurmountable, and tools have been developed in other contexts to address similar problems.

---

generally Gillian Slee & Matthew Desmond, *Eviction and Voter Turnout: The Political Consequences of Housing Instability*, 1 POL. & SOC'Y 1 (2021) (finding that “residential eviction rates negatively impacted voter turnout during the 2016 presidential election”); Allison K. Groves, Linda M. Niccolai, Danya E. Keene, Alana Rosenberg, Penelope Schlesinger, & Kim M. Blankenship, *Housing Instability and HIV Risk: Expanding our Understanding of the Impact of Eviction and Other Landlord-Related Forced Moves*, 25 AIDS & BEHAV. 1913 (2021) (finding “robust associations between landlord-related forced moves and HIV sexual risk”); Linda M. Niccolai, Kim M Blankenship, & Danya E Keene, *Eviction from Renter-Occupied Households and Rates of Sexually Transmitted Infections: A County-Level Ecological Analysis*, 15 SEXUALLY TRANSMITTED DISEASE 69 (2019) (studying the association between evictions and sexually transmitted infection rates).

Before proceeding, a note about terminology: this Article has presented a people-first approach to court data that is designed to generate insight about individuals' interactions with the courts. The unit of analysis in this approach, and the subsequent lens it applies, is that of the person rather than the institution. In the remainder of this Part, we differentiate further between data *about* people and data that *identify* particular people. Our focus is the former: the generation and/or structuring of court data to enable discovery and analysis of person-level touchpoint patterns across case types, court types, and geographies. This is an endeavor carried out in the aggregate. A single person's sequence of court touchpoints likely does not hold broader lessons for preventing collateral consequences, delivering holistic legal services, or managing problem-solving and specialized courts. However, many people's court touchpoint sequences, analyzed together, can yield important insights for advocates, researchers, court administrators, and policymakers.

Further, analyses centered on data *about* people do not require generation of or reliance on data that *identify* people. In other words, the insights available from anonymized, aggregated person-level data would not be meaningfully enriched with the addition of personally identifying information (PII) such as names or Social Security numbers. The remainder of this Part presents a set of reforms to enable the collection and assembly of court data *about* people, while also ensuring that such data do not *identify* people.

#### A. *Data Organization and Access Reforms*

The ability to engage in effective people-first research will require better, additional, and more accessible data. It will also require a different approach to thinking about court systems. There are myriad reasons why courts currently prioritize institution-first data: such data are often essential to their operation and align with the metrics most easily observed and tracked under their existing structure. Yet court data should be structured to account not only for the institutions and officials processing and deciding cases but also the people who use, depend on, and are subject to those courts.

To that end, courts could report case filing, progress, and outcomes not only on a per-case or per-court basis, but also on a per-person basis. The "person" then becomes the organizing principle and central actor across the fragmented and siloed parts of our country's multiple court systems. Doing so would provide important new sources of data, but also convey an important message: that courts are accountable not only for how they treat cases, but how they treat people. Moreover, if multiple courts and agencies adopted this model, it might be easier to make connections across entities, broadening the possibilities for research and impact. Such a change will require resources and a change in mindset, but the obstacles to achieving such an approach are not insurmountable.

What might this mean in practice? Courts could do much of this work themselves, by creating automated linking processes within their own data

collection systems based, for example, on Social Security numbers, tax identification numbers, or other unique identifying information.<sup>119</sup> It could also mean that litigants are assigned a system-wide identification (ID) number, based on Social Security information or other unique sets of identifiers and coordinated across state and federal courts at all levels. That ID number, but none of the underlying PII, could be made available to researchers or to a broader audience to enable aggregation of cases for analysis, subject to the ability to manage privacy concerns (discussed in more detail in Part IV.B). Under any of these scenarios, individual litigant names and other PII would not be released, and the data would be de-identified for external consumption as the primary research or policy interest is not in any one individual, but instead in the more general (aggregated) experience of people within the system.<sup>120</sup>

There are several examples of similar data structures already in operation. On the federal level, the Bureau of Justice Statistics has funded the creation of the Federal Justice Statistics Program (FJSP) which, under the direction of the Urban Institute, has created a dyad linking system that allows for the aggregation of records across multiple data sources and thus the ability to track person-cases through the federal criminal legal system.<sup>121</sup> FJSP's system allows researchers to merge data from multiple agencies—for example, the U.S. Marshals' Service, the Executive Office of the U.S. Attorneys, the Administrative Office of the Courts, and the U.S. Sentencing Commission. In doing so, it allows for a much more textured understanding of individuals' touchpoints across the system.<sup>122</sup> Yet even this system has its limitations. First, the existing data structure was extremely difficult to create—the report detailing its development by the Urban Institute is a daunting seventy-three pages long.<sup>123</sup> It is also limited to the criminal system and depends on the ability to create links across at least two agencies (since link-

---

<sup>119</sup> Indeed, the criminal legal system already uses fingerprints as a unique identifier. The Federal Bureau of Investigation, for example, offers a fingerprint-based "Identity History Summary Check (Rap Sheet)" for \$18 per person checked. See *How We Can Help You, Identity History Summary Checks (Rap Sheets)*, FBI, <https://www.fbi.gov/how-we-can-help-you/need-an-fbi-service-or-more-information/identity-history-summary-checks> [https://perma.cc/K5T3-ZEHK]. Thanks to Maggie Willingham for this observation.

<sup>120</sup> In the case of court-ordered expungement, both the relevant court records and an individual's system-wide identification number would need to be wiped from the various databases and master lists or repositories in which they reside. This would require that judges exercise extra care in identifying the specific types and locations of data to be expunged.

<sup>121</sup> Jessica A. Kelly, *Federal Justice Statistics Program Data Linking System*, URB. INST. (2012), <https://www.ojp.gov/library/publications/federal-justice-statistics-program-data-linking-system/> [https://perma.cc/ZPG7-8UWG].

<sup>122</sup> As in our pilot, the researchers building the dyad-linking system noted the ability to augment files with demographic information (even when information from one source was lacking such information), thus allowing for analyses that would not have been possible otherwise. See *id.* at 27. For example, by linking criminal and civil data, we were able to associate demographic characteristics with civil cases—information that is not normally available in civil court data sets.

<sup>123</sup> See *id.*



ages are created one dyad at a time). This work would likely be more difficult across civil and criminal systems, which vary in their data collection methods, and in state systems which have a higher degree of fragmentation and data siloing.

Another approach has involved linking court data and other outside data sources—for example, using person-level court data and census or commercial data to obtain demographic characteristics. Yet this approach requires complete and accurate name data, date of birth, or other unique identifiers that are often absent in civil court records and some publicly available criminal court records.<sup>124</sup> Researchers at Case Western Reserve University's Center on Urban Poverty and Community Development have successfully matched records across data sets—housing court and public assistance records—to obtain additional information about litigants (such as race) and about the effects of eviction.<sup>125</sup> That approach was limited, however, to those parties who were selected into both data sets (e.g., had been subject to an eviction filing and also applied for and obtained public assistance).

One local effort in Fulton County has attempted to create similar linkages across agencies in the hopes of identifying people in greatest need of behavioral health services. The project is based on the Sequential Intercept Model (SIM), which helps workers within the legal system divert people into community-based crisis care and treatment programs at various points of contact: arrest, court hearings, corrections, and reentry.<sup>126</sup> Key to the model is the granting of data access to all of the actors within the system across all of the affected entities, informing intercept opportunities. The effort therefore involves data sharing across multiple agencies: local jails, hospitals, state health agencies, and housing/homelessness organizations.<sup>127</sup> This kind of information sharing could be facilitated by entities that already centralize various operations across state agencies, including data management.<sup>128</sup>

---

<sup>124</sup> Rostain and O'Hara, *supra* note 22, at 496.

<sup>125</sup> Francisca García-Cobián Richter et al., *The Cleveland Eviction Study: Downstream Paths of Evictions into Homelessness and Loss of Human Capital*, REPORT BY THE CTR. ON URB. POVERTY AND CMTY. DEV. CASE WESTERN RESERVE UNIV. 8–9 (October 2019), [https://lasclv.org/wp-content/uploads/Costs-of-Eviction\\_11052019.pdf](https://lasclv.org/wp-content/uploads/Costs-of-Eviction_11052019.pdf) [<https://perma.cc/EB4C-BT9H>].

<sup>126</sup> See *The Sequential Intercept Model (SIM)*, SUBSTANCE ABUSE & MENTAL HEALTH SERVS. ADMIN. (SAMHSA), <https://www.samhsa.gov/criminal-juvenile-justice/sim-overview> [<https://perma.cc/AF54-VFMX>].

<sup>127</sup> See *Redirect People from Repeat Arrests*, TYLER TECHS. <https://www.tylertech.com/resources/case-studies/redirect-people-from-repeat-arrests> [<https://perma.cc/89C8-XBKX>].

<sup>128</sup> See, for example, Michigan's Department of Technology, Management, and Budget (DTMB), <https://www.michigan.gov/dtmb> [<https://perma.cc/S6SG-JDHS>]. See also, e.g., D. Daniel Sokol, *Technology Driven Government Law and Regulation*, 26 VA. J.L. & TECH. 1, 4–6 (2023) (proposing the creation of a “centralized data analytics unit” that would coordinate data management, analysis, and responses across government agencies).

In Fulton County, this effort is made possible through the involvement of Tyler Technologies, a leading provider of data management services to local governments and court systems. Tyler Technologies has provided the county with a platform that can be used by all of the relevant agencies to collect and clean data and then match individuals across datasets—all in an environment with “necessary controls to maintain security and work with source data” to address privacy concerns.<sup>129</sup> While this project provides a valuable model for people-first tracking through the system and real-time intervention, it is not available to researchers and highly dependent on privately-held, proprietary technology.<sup>130</sup> Moreover, given its focus on diverting people from the criminal legal system, it explores these connections only for those in the criminal court system (not civil) and thus provides only a part of the full research picture.

Despite their limitations, each of these examples offers a partial view of what restructured court data might require: data sharing across sectors, agencies, and jurisdictions; buy-in from public and private actors; and, most importantly, the creation of person-level identification schema to enable aggregation of person-level court touchpoints from across the entire legal system.

### B. *Protecting Privacy, Preserving Dignity, and Preventing Misuse*

People-first court data would enable researchers to better understand and analyze how individuals interact with various parts of the legal system across a wide variety of courts. However, with that development would come the possibility of diminished privacy and the risk of dignitary harms—the concern that data *about* people might also reveal data that *identify* people. In addition, while perhaps intended primarily for research purposes, the ability to track individuals through the court system could also be used by other parties toward their own ends—some of which may be less than desirable. Some data privacy concerns—like hacking—are not unique to people-first court data but might be elevated where people are identified across distinct data sets, making many aspects of their private lives—including information about their race, address, history of residences, and date of birth, and their familial connection to other individuals—available at once should the system be breached.

---

<sup>129</sup> *Id.*

<sup>130</sup> Indeed, the privatization of court record archives is a double-edged sword: private technology companies may offer courts the ability to build an efficient, state-of-the-art system, but at the cost of placing more and more public records under private control, with associated concerns about data access and ownership, data portability, responsibility for data security, and long-term support obligations and liabilities. *See, e.g.,* David Pozen, *Transparency's Ideological Drift*, 128 *YALE L.J.* 100, 141–44 (2018) (observing that many open data initiatives “billed as empowering the public through access to information may end up shifting the burden [and functions] of governance outside government”). Thanks to Jeff Vagle for this insight.

The concerns arising from construction of a people-first data set might be grouped into two categories: backward-looking and forward-looking. Backward-looking concerns relate to the possibility of revealing, and possibly allowing misuse of, extensive information about an individual and their past history, including details about a variety of interactions they may have had with the legal system. Forward-looking concerns relate to the likelihood that more comprehensive information about an individual's past might be used to make predictions about that individual's future—or that generalized information about the experience of a certain group of people might be used to make conclusions about how members of that group will act in the future. These concerns are addressed in turn below.

### 1. *Backward-Looking Concerns*

One set of backward-looking concerns stems from the fear that existing data might be exploited or misused, based on the specific ways in which a people-first approach restructures or reports those data. For example, creating additional linkages across data sets might make personal information more easily available to a broader audience and could be misused in ways that are harmful to individuals.

The concern here is not that court data are themselves private and would be revealed by a people-first approach. After all, most criminal prosecutions and civil lawsuits generate public records that anyone can view at any courthouse clerk's office or online, where electronic court records are available. However, the fragmentation and siloing of courts and their records act as a *de facto* information shield. As noted above, reorganizing courts' data around the person risks revealing previously existent, but buried, information. One can imagine, for example, that linked court records might enable identity theft by exposing a person's address history and financial information. Likewise, some court data transparency advocates fear that abusers would track domestic violence survivors by following their trail through the courts.<sup>131</sup>

To avoid these developments, necessary connections between and among individuals could be made by the courts themselves on the back end, through the creation of an independent, de-personified identifier that is used across systems—so that personal information need not be used as a surface-level means for linking individual data.<sup>132</sup> Such a method would actually be

---

<sup>131</sup> See, e.g., *Domestic Violence and Privacy*, ELEC. PRIV. INFO. CTR., <https://archive.epic.org/privacy/dv/> [<https://perma.cc/EQ8M-3G7V>] (raising concerns that domestic violence survivors would be exposed to stalking as a result of “[d]ata brokers min[ing] these records to resell and build profiles on individuals”).

<sup>132</sup> There may be concerns about people-first data being used not only externally, but also internally—for example, a clerk or judge using information from an unrelated proceeding prejudicially. Such a possibility certainly already exists in the current system, as it is likely less difficult for court personnel to obtain such data, but could also be avoided through restricted (internal) access to the independent identifiers.

more privacy-protective than the system we used in our pilot study, which took public-facing court records—names, addresses, and all—and constructed linked records for analysis. Our proposal would obviate the need for any use of names and other PII and the associated privacy risks.

Access to and use of such data can also be heavily restricted. For example, the Criminal Justice Administrative Records System (CJARS), a “nationally integrated repository of data to following individuals through the criminal justice system” housed at the University of Michigan,<sup>133</sup> anonymizes its data and creates a unique identifier for each individual.<sup>134</sup> In addition, CJARS data groupings that are smaller than a congressional district cannot be published or exported from the secure online environment<sup>135</sup> and access to the data is granted to researchers on a restricted-use/approval basis managed by the U.S. Census Bureau.<sup>136</sup> Although this level of aggregation means that conclusions can only be made about groups of people, it also protects the privacy of individual data subjects.

Similarly, data from the Federal Justice Statistics Program discussed above is only made available to other users on a restricted-use/approval basis, through the Inter-university Consortium of Political and Social Research (ICPSR) at the University of Michigan.<sup>137</sup> Because these types of systems are, at their core, based on individual records and existing linkages among those records, they leave open the possibility—however remote, given heightened data security protocols and restricted access—of systemic hacking and leaking of sensitive information.

Some researchers have tried to address this issue through the use of “differential privacy”—a set of algorithmic tools that allow researchers to understand the characteristics and behaviors of individual experiences in the aggregate without the analysis turning on or identifying any one individual.<sup>138</sup> Differential privacy tools do more than just anonymize personal information, which others have suggested is insufficient to protect individual privacy.<sup>139</sup> Rather than merely inputting the information of private individuals and using it to generate aggregate data, which still relies on that individual information, differential privacy adds “random noise,” also known as “statistical noise,” to the analysis to produce approximate outcomes, masking the original data before it is used for analysis and eliminating the possi-

---

<sup>133</sup> *Criminal Justice Administrative Records System*, UNIV. OF MICH., <https://cjars.isr.umich.edu/> [<https://perma.cc/6UVW-REBB>].

<sup>134</sup> See Jason Tashea, *Justice-as-a-Platform*, MIT COMPUTATIONAL LAW REP., <https://law.mit.edu/pub/justiceasaplatform/release/3> [<https://perma.cc/YU5F-474R>].

<sup>135</sup> See Tashea, *supra* note 134, at 11.

<sup>136</sup> See *Criminal Justice Administrative Records System*, *supra* note 133.

<sup>137</sup> See Kelly, *supra* note 121.

<sup>138</sup> Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David R. O’Brien, & Salil Vadhan, *Differential Privacy: A Primer for a Non-technical Audience*, THE PRIVACY TOOLS FOR SHARING RSCH. DATA PROJECT AT HARV. UNIV. 3-10 (2018) [https://privacytools.seas.harvard.edu/files/privacytools/files/pedagogical-document-dp\\_new.pdf](https://privacytools.seas.harvard.edu/files/privacytools/files/pedagogical-document-dp_new.pdf) [<https://perma.cc/XX28-ZNKZ>].

<sup>139</sup> Tashea, *supra* note 134, at 11.

bility that true individual data can be imputed from the aggregate results.<sup>140</sup> Because a people-first approach to court data focuses on the experience of aggregated groups of people rather than identifiable individuals, differential privacy tools represent an alternative mechanism for allowing researchers to obtain the benefits of data *about* people without the risk of creating data that *identify* people.<sup>141</sup>

## 2. Forward-Looking Concerns

Another set of concerns regarding people-first data relates not to what inquiring minds might be able to reveal about a particular individual's past or private information, but instead to make predictions about an individual's future behavior. The rise of big data and artificial intelligence has come with increased concern about how the use of these tools by governmental entities can render all people, particularly the poor, vulnerable to privacy harms, including data exposure and misuse, and increased surveillance and punishment.<sup>142</sup>

While these concerns are already implicated by a number of existing public data repositories where massive amounts of information are collected,<sup>143</sup> creating additional linkages across existing data sets would arguably exacerbate the possibilities for harm. Consider, for example, predictive policing, a law enforcement strategy that uses historical crime data and other information to identify locations where criminal activity is likely to occur.<sup>144</sup> Critics of this use of technology have explained how it is likely to unfairly target individuals on the basis of race and class while providing officers with the seemingly objective cover of data-driven analysis.<sup>145</sup> Many existing predictive policing models are place-based;<sup>146</sup> the promulgation of people-first data could, in theory, provide a way to base predictions on patterns of individual behavior as well as geographic frequency. In its most dangerous form,

<sup>140</sup> Nissim et al., *supra* note 138; *see also, e.g., Apple Differential Privacy Technical Overview*, APPLE (July 28, 2022), [https://www.apple.com/privacy/docs/Differential\\_Privacy\\_Overview.pdf](https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf) [<https://perma.cc/U3EU-ETHU>].

<sup>141</sup> *See generally* Cynthia Dwork & Aaron Roth, *The Algorithmic Foundations of Differential Privacy*, 9 *FOUND. & TRENDS IN THEORETICAL COMPUT. SCI.* 211 (2014).

<sup>142</sup> *See generally* VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2018); Elizabeth E. Joh, *The Unexpected Consequences of Automation in Policing*, 75 *S.M.U. L. REV.* 507, 518, 528 (2022); Mary Madden, Michele Gilman, Karen Levy, & Alice Marwick, *Privacy, Poverty and Big Data: A Matrix of Vulnerabilities for Poor Americans*, 95 *WASH. U. L. REV.* 53, 55 (2017).

<sup>143</sup> *See, e.g.,* Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 *CORNELL L. REV.* 1875, 1892-1905 (2020) (providing multiple examples of how large bodies of aggregated public data are subject to possible data breaches and predictive analysis by machine learning tools, raising privacy concerns).

<sup>144</sup> *See* Andrew Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 *EMORY L. J.* 259, 265 (2012).

<sup>145</sup> *See* Joh, *supra* note 141, at 3; Ferguson, *supra* note 144, at 265, 307-08.

<sup>146</sup> *See id.*, at 266, 274 (explaining that place-based policing approaches focus on where a crime might be committed or vulnerabilities that make particular locations susceptible to criminal activity).

it could produce fodder for predictions not only of where crime will occur, but also who will engage in it.

Predictive policing can be used to help law enforcement allocate their time, attention, and resources.<sup>147</sup> One could imagine that in the court context, people-first data could be used in a similar fashion by courts, related government institutions, and service providers to target certain individuals for differential treatment. Though such targeting by law enforcement may lead people into further entanglement with the criminal legal system, there may be more benign outcomes associated with identifying when and how individuals interact with the courts. Some of these uses could also be helpful—for example, by flagging individuals who may have concurrent or sequentially-close legal matters that could be addressed in tandem by a service provider or court.

Of course, there is also the possibility that such data could be used for more harmful purposes, particularly if made available to private entities or if there is a market for private companies to profit from its exploitation. Consider the data analytics firm Palantir, which combines and analyzes personal data acquired from commercial, proprietary, and public data sets to identify trends and relationships that can be used for predictive purposes.<sup>148</sup> Palantir's technology has been widely used by government entities and law enforcement; evidence suggests that its system has helped, among other tasks, to identify and deport family members of unaccompanied minors.<sup>149</sup>

In the court context, people-first data could be used by private entities to blacklist individuals who have certain patterns of court touchpoints from obtaining available goods, services, or resources, including employment or housing, on the theory that their past court involvement predicts continued future involvement.<sup>150</sup> Further, judges or prosecutors who exercise discretion in making charging decisions, imposing sentences, and deciding damages might be prejudiced against litigants if they could easily access data on litigants' full panoply of court touchpoints.

As one of us has explained in other work, this use of person-first court data for predictive purposes risks both dignitary and economic harm:

---

<sup>147</sup> See Madden et al., *supra* note 142, at 106.

<sup>148</sup> See Tamar Hoffman, *Debt and Policing: The Case to Abolish Credit Surveillance*, 29 GEO. J. ON POVERTY L. & POL'Y 93, 111-13 (2021); Peter Waldman, Lizette Chapman & Jordan Robertson, *Palantir Knows Everything About You*, BLOOMBERG BUSINESSWEEK, (Apr. 19, 2018) <https://www.bloomberg.com/features/2018-palantir-peter-thiel/> [<https://perma.cc/V692-N6MR>].

<sup>149</sup> See Kate Crawford & Jason Schultz, *AI Systems as State Actors*, 119 COLUM. L. REV. 1941, 1946 (2019).

<sup>150</sup> See, e.g., Kathryn A. Sabbeth, *(Under)enforcement of Poor Tenants' Rights*, 27 GEO. J. POVERTY L. & POL'Y. 97, 109 (explaining how eviction filings can place tenants on a "blacklist" used by landlords to screen out applicants); see also Siya U. Hegde, *I Am Not a Nuisance: Decriminalizing Domestic Violence Across New York's Civil Housing and Criminal Justice Systems*, 29 GEO. J. POVERTY L. & POL'Y 1 (2022) (describing how domestic violence victims and survivors, as well as alleged abusers, may face eviction as a result of their domestic violence involvement).

[Such practices] can reduce people to their worst past acts and prevent them from changing course. . . . By reducing people to feature sets and exploiting the features that are most predictive of outcomes, but perhaps least representative of people's full selves, computational tools [and the data on which they operate] enact dignitary harm. . . , reducing social mobility and locking people into place.<sup>151</sup>

One way to cabin such harmful uses may be to structure the data such that details on any one individual are not accessible, but instead reported only in the aggregate—and to restrict its availability to those engaged in relevant research, as described above. While the former approach would not prevent prediction based on an individual's association with a larger group-defining characteristic, such as race, gender, or certain case patterns, it would eliminate the possibility of direct individual surveillance. Further, algorithmic approaches to shielding identity such as the differential privacy techniques described above would block individual-level predictions and their associated risks and harms. Moreover, in the case of prosecutorial or judicial decision-making, charging standards and sentencing guidelines could be revised to detail permissive and prohibited uses of aggregated court touchpoint data as it pertains to individuals' cases.<sup>152</sup>

On the pertinence of aggregated court touchpoint data to individual cases, we acknowledge a possible tension with our discussion in Part III.B above about the potential benefits of person-first court data for projects like holistic legal services. For example, if person-first data provide a useful and beneficial picture of the full range of a client's interactions with the legal system, why not give a legal services provider access to data that *identify* that client? The answer is that the risks of generating and making non-anonymized data broadly available outweigh the potential benefits detailed in this and previous sections. In addition, legal services lawyers and their clients—like prosecutors and judges—have alternative ways to assemble their own individual, identifiable court touchpoints data: clients' own report-

---

<sup>151</sup> Charlotte S. Alexander, *Litigation Outcome Prediction, Access to Justice, and Legal Endogeneity*, in *LEGAL TECH AND THE FUTURE OF CIVIL JUSTICE* (David Freeman ed., 2023).

<sup>152</sup> Similar precautionary measures have been recommended in the United Kingdom, where open justice policies require that tribunal judgments be made publicly accessible. See Suja A. Thomas, *Can the US Learn from Open Justice in UK Employment Tribunals*, JOTWELL (Feb. 24, 2023), <https://courtslaw.jotwell.com/can-the-us-learn-from-open-justice-in-uk-employment-tribunals/> [<https://perma.cc/7XEK-AVHQ>]. Based on fears about how such information might be used by employers and by artificial intelligence, as well as in future settlement negotiations, scholars have recently recommended anonymization of the data (through omission of party names) and legally prohibiting employers from denying employment based on an employee's litigation history. See Zoe Adams, Abi Adams-Prassl & Jeremias Adams-Prassl, *Online Tribunal Judgments and the Limits of Open Justice*, 42 *LEGAL STUD.* 42 (2022).

ing, plus public records searches.<sup>153</sup> Once assembled, legal services providers could match their clients' sequence of court touchpoints to aggregated, deidentified data sources such as those produced in this pilot project, to gain insight about common patterns and outcomes experienced by other individuals who have interacted with the courts. Thus, the distinction between data *about* people and data that *identify* people remains intact.

In the end, we suggest that the potential misuse of people-first data counsels not towards preventing its creation, but instead towards heightened critique, sensitivity, and thoughtfulness around identifying such harms and minimizing or eliminating them during the data's use and development. Here, we follow Erin Collins' call to "heed the advice of QuantCrit scholars and embrace an attitude of 'principled ambivalence' towards the quantitative data we do have."<sup>154</sup> Researchers and policymakers who seek to create and use people-first data should not become uncritical cheerleaders but should instead proceed with eyes open to potential benefits and potential harms.

## CONCLUSION

The process of restructuring existing court data to generate people-first data—or creating systems to generate such data anew—can be complex and raises a number of significant privacy and (mis)use concerns. Yet there is also incredible promise in data that can help those engaged with the court system from a use, management, or research perspective to understand how people, and not courts, experience such systems. People-first data can add a critical quantitative analytic component to the growing body of research exploring how people understand the legal system.<sup>155</sup> In addition, a people-first approach can provide important new data for areas of research focused on how different parts of the legal system—various levels and types of courts, sentencing, and legal representation—relate to one another through the individual's touchpoints over time. Aside from its value to academic research, people-first data are essential to court and legal reform, including aiding the

---

<sup>153</sup> Indeed, this approach preserves clients' dignity, as they have control over their own story and telling of their history and can collaborate with their attorney in collecting and verifying data about their past. Thanks to Nate Vogel for this insight.

<sup>154</sup> Collins, *supra* note 11. QuantCrit scholars, in turn, suggest that "numbers are not neutral," and that "voice and insight are vital: data cannot 'speak for itself' and critical analyses should be informed by the experiential knowledge of marginalized groups." David Gillborn, Paul Warmington, & Sean Demack, *QuantCrit: Education, Policy, 'Big Data' and Principles for a Critical Race Theory of Statistics*, 21 RACE, ETHNICITY & EDUC. 158, 158 (2017).

<sup>155</sup> For example, some research has already made headway in understanding individuals' experiences across the civil-criminal divide. Yet this largely qualitative research could be incredibly enriched by the data described herein. See, e.g., Lauren Sudeall & Ruth Richardson, *Unfamiliar Justice: Indigent Criminal Defendants' Experiences with Civil Legal Needs*, 52 U.C. DAVIS L. REV. 2105 (2019) (exploring how indigent criminal defendants understand, experience, and respond to civil legal problems); Greene, *supra* note 26 (describing how past negative experiences with the criminal justice system inform individuals' decisions to seek help with civil legal problems).



determination of where and how certain services and interventions can improve outcomes.

In our discussion of people-first data, the methodology for developing it, and our pilot project, we have highlighted a number of obstacles, caveats, and concerns. Yet, we hope this Article is only the beginning of a larger conversation. We invite academics, researchers, policymakers, court administrators, and other stakeholders—and of course the litigants who interact with courts every day—to join this discussion and contribute their own ideas for what data are needed, how they can best be used, and how to safeguard the people and interests involved. Just like the data that are the subject of this Article, the discussion of how to generate and when and how to use people-first data should not be a top-down endeavor but informed by those who have the most at stake.

## APPENDIX A

TABLE 8: CROSSWALK AMONG ORIGINAL CASE TYPE ASSIGNED BY COURTS AND STANDARDIZED GRANULAR AND CONSOLIDATED CASE TYPES ASSIGNED BY RESEARCHERS

<i>Original Case Type</i>	<i>Standardized Case Type (granular)</i>	<i>Standardized Case Type (consolidated)</i>
Abandoned Motor Vehicle	Abandoned Motor Vehicle	Other
ACCOUNT	Account/Contract	Other
APPEAL	Appeal	Other
APPEAL FROM MAGISTRATE COURT	Appeal	Other
AUTO ACCIDENT	Tort	Other
AUTOMOBILE TORT	Tort	Other
CF-CAPITAL FELONY	Capital Felony	Felony
CHILD SUPPORT	Child Support	Family
CHILD SUPPORT/PRIVATE ATTORNEY	Child Support	Family
CONDEMNATION OF MONEY/AUTOMOBILE	Other Civil Cause of Action	Other
CONTINUING GARNISHMENT	Garnishment	Other
CONTRACT	Account/Contract	Other
CONTRACT/ACCOUNT	Account/Contract	Other
Conversion	Tort	Other
Conversion Civil Case	Tort	Other
CS-CAPITAL SEX CRIMES	Capital Sex Crimes	Felony
CUSTODY	Custody	Family
DAMAGES	Other Civil Cause of Action	Other
Dispossessory	Dispossessory	Dispossessory
DISPOSSESSORY-STATE COURT	Dispossessory	Dispossessory
Distress Warrant	Dispossessory	Dispossessory
DIVORCE - CONTESTED	Divorce	Family
DIVORCE - UNCONTESTED	Divorce	Family
DIVORCE BY PUBLICATION	Divorce	Family
E-FILED	Other Civil Cause of Action	Other

<i>Original Case Type</i>	<i>Standardized Case Type (granular)</i>	<i>Standardized Case Type (consolidated)</i>
EFILEDSTATE CIVIL	Other Civil Cause of Action	Other
FOREIGN JUDGMENT	Other Civil Cause of Action	Other
GC-GENERAL COMPLEX	General Complex	Felony
JUDICIAL REVIEW	Other Civil Cause of Action	Other
LEGITIMATION	Legitimation	Family
Magistrate Account	Account/Contract	Other
Magistrate Contract	Account/Contract	Other
Magistrate Note	Other Civil Cause of Action	Other
Magistrate Tort	Tort	Other
Magistrate Trover	Tort	Other
MC-MURDER (1-3 DFTS)	Murder	Felony
MEDICAL MALPRACTICE	Tort	Other
MISDEMEANOR	Misdemeanor	Misdemeanor
MM-MULTIPLE MURDER (4+ DFTS)	Murder	Felony
MODIFICATION	Other Civil Cause of Action	Other
NAME CHANGE PETITION	Other Civil Cause of Action	Other
NF-NON CAPITAL FELONY	Non-Capital Felony	Felony
NJ-NON COMPLEX	Non-Complex	Felony
NOTE	Other Civil Cause of Action	Other
OTHER CIVIL CAUSE OF ACTION	Other Civil Cause of Action	Other
PERSONAL INJURY	Tort	Other
QUIET TITLE	Other Civil Cause of Action	Other
REGULAR GARNISHMENT	Garnishment	Other
RN-PREFILE / CRIMINAL COMPLAINT	Prefile / Criminal Complaint	Profile
RO-REPEAT OFFENDER-NON COMPLEX	Repeat Offender-Non Complex	Felony

<i>Original Case Type</i>	<i>Standardized Case Type (granular)</i>	<i>Standardized Case Type (consolidated)</i>
Small Claims	Small Claims	Other
State Court Dispossessory	Dispossessory	Dispossessory
TORT	Tort	Other
TORT/NEGLIGENCE	Tort	Other
Traffic	Traffic	Traffic
TRANSFER - CIVIL	Other Civil Cause of Action	Other
TRUST	Other Civil Cause of Action	Other

## APPENDIX B

TABLE 9: CONSOLIDATED CASE TYPES WITH DISTRIBUTION OF CASE STATUS PER CASE TYPE

<i>Case Status</i>	<i>Dispos- sessory</i>	<i>Family</i>	<i>Felony</i>	<i>Misde- meanor</i>	<i>Other</i>	<i>Prefile</i>	<i>Traffic</i>
Closed	70%	91%	92%	82%	72%	56%	49%
Open	30%	9%	4%	8%	27%	40%	41%
Dormant/hold/ stayed/inactive	0%	0%	4%	8%	1%	0%	1%
Paid	0%	0%	0%	0%	0%	0%	5%
Transfer	0%	0%	0%	0%	0%	4%	0%
Forwarded to solicitor general	0%	0%	0%	0%	0%	0%	2%
Bound over to State Court	0%	0%	0%	2%	0%	0%	0%
Withdrawn	0%	0%	0%	0%	0%	0%	1%
Time to pay	0%	0%	0%	0%	0%	0%	1%